University of Alberta

Library Release Form

Name of Author: Chun-Fu Liu

Title of Thesis: Completeness of Case Ascertainment in the Alberta Cancer Registry

Degree: Master of Science

Year this Degree Granted: 2001

UNIVERSITY OF ALBERTA

Completeness of Case Ascertainment in the Alberta Cancer Registry

by

Chun-Fu Liu ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

MEDICAL SCIENCES – PUBLIC HEALTH SCIENCES

EDMONTON, ALBERTA

FALL 2001

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *Completeness of Case Ascertainment in the Alberta Cancer Registry* submitted by Chun-Fu Liu in partial fulfillment of the requirements for the degree of Master of Science in Medical Sciences – Public Health Sciences.

ABSTRACT

Registration completeness in the Alberta Cancer Registry for the years 1994-96 was determined. A major deterministic record linkage between Registry data and hospital discharge and day procedure data from Alberta Health and Wellness for the years 1994-1996 was undertaken. Also, Incidence:Mortality ratios, indicators of percent Histological Verification, and percent Death Certificate Only measures were calculated. A case re-finding at the University of Alberta Hospitals was conducted.

Registry case coverage estimated from the major linkage is 93.0% (92.8%-93.2%). Among the main cancer sites investigated, female breast (98.3%; 97.9%-98.7%) and leukemia (83.0%; 80.7%-85.4%) had the highest and lowest case coverage respectively.

Factors contributing to incompleteness were assessed. Results were stratified by demographic and geographic factors, illustrating variations attributable to Registry completeness. Results from the other methods were compared to the major linkage. Findings from this study will help improve cancer registration in Alberta for better cancer control.

ACKNOWLEDGEMENTS

Many people contributed to the success of this thesis. I acknowledge certain individuals specifically.

My thesis committee comprising Dr. Colin L. Soskolne, Dr. Juanita Hatcher and Dr. Anthony L.A. Fields provided constant support. I am especially grateful for the patient guidance of my co-supervisors, Dr. Colin L. Soskolne and Dr. Juanita Hatcher. Mr. Dave Clyburn in the Academic Support Centre provided thoughtful service in editing.

Members of the Division of Epidemiology, Prevention, and Screening in the Alberta Cancer Board provided ongoing technical support. In particular, I would like to thank Ms. Herta Gaedke, Mr. Voon Siaw, Ms. Maxine Raphael, and Mr. John Hanson for their assistance and spiritual encouragement.

I gratefully acknowledge the financial support received from Alberta Cancer Board.

Finally, I would like to thank my family, including my parents and brothers, for their continued support and belief in my potential. I am especially grateful for all the support and encouragement from my friend, Poshen Wang.

TABLE OF CONTENTS

Signature Page

Abstract

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Chapter

I. Introduction 1

II. Literature Review 5

- A. Population-based Cancer Registries 5
- B. The Role of Registries in Cancer Control 6
- C. Data Quality in Cancer Registries 7
- D. Methods of Estimating Registration Completeness 9
  - 1. Data Sources 9
    - 1.1 Number of Sources/Notifications Per Case 10
    - 1.2 The Death Certificate Method 10
    - 1.3 The Percent Histological Verification of Diagnosis (HV%) Method 11
  - 2. Independent Case Ascertainment 12
    - 2.1 Comparison with Independent Data Sets 12
      - 2.1.1 Record linkage 12
    - 2.2 Re-screening of Cases 14
    - 2.3 The Capture/Recapture Methods 15
    - 2.4 The Incidence:Mortality (I:M) Ratio 15
  - 3. The Historic Data Method 18
    - 3.1 Incidence Rates over Time 18

3.2 Comparison of Incidence in Different Populations 19
3.3 Age-specific Incidence Curves 19
4. A Statistical Modeling Technique 19
E. Cancer Registries in Canada 20
F. The Alberta Cancer Registry (ACR) 22
G. The Effect of Different Coding Rules on Estimates of Registry Completeness 23
H. Confidentiality and Ethics 27
I. Legislation on Accessing Cancer Information 31
J. Methods Appropriate to Estimate Alberta Cancer Registry Completeness 33
1. Comparisons between AHW and ACR Data 33
2. Use of Incidence:Mortality (I:M) Ratios 33
3. Use of Indicators of Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%) 34
4. Case Re-finding in the University of Alberta Hospitals (UAH) 34
III. Materials 36
A. The Alberta Cancer Registry 36
1. The AHW-ACR Record Linkage and UAH Case Re-finding 40
2. The Incidence:Mortality (I:M) Ratio 40
3. Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%) 41
B. Alberta Health and Wellness Data 41
C. University of Alberta Hospitals Data 43
D. U.S. SEER Incidence and NMDB Mortality Data 44
1. SEER Incidence 44
2. NMDB Mortality 44

IV. Methods 45

A. The AHW-ACR Record Linkage 45

B. The Incidence:Mortality (I:M) Ratio 52

1. ACR Completeness Estimates 52

2. Impacts of Different Rules on Completeness Estimates 54

C. Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%) 54

D. Case Re-finding in the University of Alberta Hospitals 55

E. Stratification by Demographic and Geographic Factors 59

F. Comparison of Methods 60

V. Results 61

A. The AHW-ACR Record Linkage 61

B. The Incidence:Mortality (I:M) Ratio 66

1. ACR Completeness Estimates 66

2. Impacts of Different Rules on Completeness Estimates 70

C. Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%) as Completeness Indicators 74

D. The UAH Case Re-finding 77

E. Comparison of Methods 83

F. Summary 88

VI. Discussion 89

A. ACR Completeness Estimates 89

1. The AHW-ACR Record Linkage 89

2. The Incidence:Mortality (I:M) Ratio Method 91

2.1 ACR Completeness Estimates 91

2.2 Impacts of Different Rules on Completeness Estimates 93

3. Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%) as Completeness Indicators 93

4. Case Re-finding in the UAH 94

- B. Stratification by Demographic and Geographic Factors 95
  - 1. Sex Stratification 95
  - 2. Age Group Stratification 96
  - 3. Regional Stratification 96
- C. Comparison of Methods 97
- D. Strengths and Limitations of the Current Study 98
  - 1. Strengths 98
  - 2. Limitations 99

VII. Conclusions and Recommendations 102

References 104

LIST OF TABLES

Table 2.1 ICD-0-2 Topology Codes to Be Considered One Primary Site When Determining Multiple Primaries for Both ACR & IARC Coding 25

Table 2.2 Groups of Malignant Tumours Considered to Be Histologically Different for the Purpose of Defining Multiple Tumours Using IARC Rules 26

Table 4.1 Codes in Two Files Considered Equivalent 51

Table 4.2 Comparisons of Crude and Standardized I:M Ratio in the ACR, 1994-96 53

Table 5.1 Results of the AHW-ACR Record Linkage, 1994-96 62

Table 5.2 Results of the Case Follow-back for the AHW-ACR Record Linkage 63

Table 5.3 Results of the AHW-ACR Record Linkage Stratified by Sex, Age Group, and Region, 1994-96 65

Table 5.4 Crude I:M Ratios for ACR Completeness by Year for 1994-96 68

Table 5.5 Crude I:M Ratios for ACR Completeness by Sex, Age Group, and Region, 1994-96 69

Table 5.6 The Differences in Incidence and Crude I:M Ratio as the Result of the Adoption of Different Coding Rules in the ACR, 1994-96 72

Table 5.7 Comparison of ACR Completeness Estimates Using I:M Ratios between the Adoption of SEER and IARC Rules, 1994-96 73

Table 5.8 HV% and DCO% as ACR Completeness Indicators by Year, 1994-96 75

Table 5.9 HV% and DCO% as ACR Completeness Indicators by Sex, Age Group, and Region, 1994-96 76

Table 5.10 Results of the UAH Case Re-finding, 1994-96 79

Table 5.11 Results of the UAH Case Re-finding Stratified by Sex, Age Group, and Region, 1994-96 82

Table 5.12 Comparisons of ACR Completeness Estimates by Different Methods with Sex Stratification, 1994-96 85

Table 5.13 Comparisons of ACR Completeness Estimates by Different Methods with Regional Stratification, 1994-96 87

LIST OF FIGURES

Figure 4.1 Alberta Health and Wellness (AHW) and Alberta Cancer Registry (ACR) Record Linkage 50

Figure 4.2 University of Alberta Hospitals (UAH) Case Re-finding 58

LIST OF ABBREVIATIONS

ACR	Alberta Cancer Registry
AHCIP	Alberta Health Care Insurance Plan
AHW	Alberta Health and Wellness
CCR	Canadian Cancer Registry
CCOCS	Canadian Coalition on Cancer Surveillance
CHA	Capital Health Authority
CINA	Cancer in North America
DCN	Death Certificate Notification
DCO	Death Certificate Only
FOIPP	Freedom of Information and Protection of Privacy
HV	Histological Verification
IACR	International Association of Cancer Registries
IARC	International Agency for Research on Cancer
ICD-O	International Classification of Diseases for Oncology
I:M Ratio	Incidence:Mortality Ratio
NAACCR	North American Association of Central Cancer Registries
NCI	National Cancer Institute
NMDB	National Mortality Data Base
PHN	Personal Health Number
SEER	Surveillance, Epidemiology, and End Results Program
UAH	University of Alberta Hospitals

CHAPTER I. INTRODUCTION

While researchers around the world are conducting epidemiological studies based on cancer registry data, the quality of data from these registries may not have been systematically evaluated. As a consequence, the results of such epidemiological studies may be questionable. The quality of data in terms of completeness, timeliness, and accuracy are of great concern when investigating potential risk factors for the causes of cancers. In addition, the roles in surveillance and prevention that cancer registries play are undermined when the data quality in cancer registries is not assured. The results produced if cancer registries have poor quality data could have misleading impacts in planning for cancer control. Therefore, the assessment of data quality in cancer registries becomes a priority for cancer control.

Data quality varies in cancer registries around the world. In Canada, cancer registries exist in all provinces and territories, providing complete coverage nationally, but variations in terms of completeness, accuracy, and timeliness across the country cannot be ignored. Currently, we do not have a standard approach for assessing the data quality of cancer registries in Canada. The Canadian Coalition on Cancer Surveillance (CCOCS) has addressed the importance of putting a cancer surveillance prototype into place. A national pilot study to determine a cost-effective, consistent, and flexible system to be used for evaluating the completeness of registration in cancer registries across the country is currently being undertaken. Results and findings from this study in Alberta will contribute to the national study.

Completeness of registration may be defined as the extent to which all the incident cancers occurring in a target population are included in the registry. Various methods can be used to estimate registration completeness, including the use of data sources, independent case ascertainment, and historic data methods. Consideration should be given to the different case-finding mechanisms among registries as well as to the costs and resources associated with the methods chosen, when implementing case ascertainment activities.

The specific objectives of this study are:

1. To determine the completeness of the ACR for the years 1994-96
2. To identify factors that may affect ACR completeness
3. To compare methods for assessing registration completeness

The hypotheses of this study include:

1. Registration completeness may be affected by factors such as the primary cancer site, and demographic and geographic factors.
2. Methods vary in their ability and ease to measure completeness.

One of the methods used in this study was to compare the data in the Alberta Cancer Registry (ACR) and other independent data maintained by Alberta Health and Wellness (AHW), including hospital discharge and day surgery files for the years of 1994 to 1996. A deterministic record linkage was used to link the two data sets. Because all Alberta residents are covered by the Alberta Health Care Insurance Plan (AHCIP)

provided by AHW, information on all services provided to cancer patients in Alberta is available from AHW. Although all sites were investigated, cancers of the female breast, prostate, colon-rectum, and lung were of particular interest because of their high incidence in Alberta. Leukemia and melanoma of the skin were investigated because of problems suspected in their reporting. In addition, pancreatic cancer was investigated owing to the short survival after diagnosis.

Those cases identified in the AHW data that were not registered in the ACR, as well as those cases for which the ACR primary cancer site did not agree with that recorded as a discharge diagnosis on AHW data, were followed back by ACR staff to confirm the diagnosis. Registration completeness was estimated by levels of case coverage, which was defined as 1 – [unlinked AHW cancers/(number of ACR cancers + unlinked AHW cancers)], and was derived based on the percentage of diagnosed cancers in the AHW data file that the ACR had registered.

Other methods examined were the use of independent data sources for the calculation of the Incidence:Mortality (I:M) ratio, and indicators such as percent Histological Verification (HV%) and percent Death Certificate Only (DCO%). Completeness of case coverage may be estimated by comparing of the age-standardized Alberta I:M ratio to a gold standard I:M ratio. The effect of different standard populations and different coding rules for multiple primaries was examined.

In order to supplement the methods mentioned above, a review of computerized outpatient medical records for the years 1994 to 1996 at the University of Alberta Hospitals (UAH) in Edmonton was conducted for this case re-finding exercise. Health care utilization information is centralized and electronically recorded in the Capital

Health region, making this feasible. The cases from the UAH were matched to cases that already are in the ACR to identify unreported cases. Cases from the UAH data that did not have records on the ACR database, as well as cases from the UAH data when diagnosis did not agree with that recorded on the ACR, were followed back to confirm the diagnosis.

To investigate the factors contributing to completeness and to compare different methods used in estimating completeness, stratified analyses based on attributable factors were done for each of the methods used. Patterns observed in the comparison can better explain registration completeness, and the most cost-efficient and appropriate method for ACR completeness estimates can be determined.

In summary, completeness of case ascertainment in the ACR for the years of 1994-96 was estimated by each of several methods. Identification of factors contributing to registration completeness to improve data quality in cancer registries was achieved. An optimal configuration of methods for ACR completeness estimates is suggested. The contribution of cancer registry data to epidemiological studies would thereby be enhanced, and the registry function of surveillance, prevention, early detection, and even treatment and care would be improved to better support cancer control. Methods and results from this study can be generalized and thus could be incorporated into studies of other registries in other centres.

CHAPTER II. LITERATURE REVIEW

Cancer registries are crucial in the support of cancer control, but the function of cancer registries may be compromised if the data are of poor quality. This chapter reviews the literature on population-based cancer registries, on the roles registries play in cancer control, on the importance of having good quality data in cancer registries, and information on methods used to estimate registration completeness. Furthermore, descriptions of the status of cancer registries in Canada and in Alberta are given. Methods appropriate for estimating ACR completeness are discussed with consideration to the implications for the different coding rules used among registries. Issues concerning confidentiality and ethics pertaining to accessing cancer information as well as legislation associated with assessing cancer information are discussed.

A. Population-based Cancer Registries

Cancer registration is usually defined in the literature as the processes involved in the continued systematic collection of data on the occurrence and characteristics of reportable cancers [1-4]. In paper or electronic format, information about individuals with cancers is collected by registries to support cancer control activities [5]. Through the collection of information on cancer cases and the assessment of the impact of intervention programs, the main objectives of reducing cancer burdens in the population for a cancer registry are achieved [3]. There are two main types of cancer registries -- population-based and hospital (non-population-based) registries.

Population-based registries include all cancer cases diagnosed in a defined geographic area and are therefore a type of registry that allows for the calculation of incidence and mortality rates [4]. Population-based registries often contain composite data from more than one reporting facility such as a hospital or a clinic. Hospital-based registries cover only those cases diagnosed and treated in certain hospitals, and do not provide full coverage of all the cancer cases diagnosed in a geographic area. Thus, the role that population-based cancer registries play in support of cancer control is crucial.

B. The Role of Registries in Cancer Control

The roles of cancer registries in cancer control are gathering information about cancer in a community, providing the data needed to elucidate the causes of the many different cancers, and evaluating the programs for controlling them [2]. In general, the activities involved in cancer control through the utilization of cancer registry data include the following [6,7]:

1. the continued assessment of the cancer burden in the population
2. the provision of data for epidemiological studies to investigate the risk factors associated with cancers
3. the evaluation and planning of patient care services
4. the evaluation of early diagnosis and treatment programs

Cancer registry information may be used for various purposes, and the value of the data increases if comparability over time is maintained [3]. The fundamental surveillance

activity of cancer registries is to monitor temporal trends and geographic patterns with respect to incidence, mortality, survival, and health services utilization. Cancer registration data may be used to measure incidence and survival according to other characteristics such as age, sex, and socio-economic status [8]. Ideally, better health planning and resource allocation can be achieved through analyzing and interpreting the data from cancer registries for prevention, screening, diagnosis, treatment, rehabilitation, and palliative care. Cancer registries are also useful for identifying subjects in case-control studies, as well as for efficiently ascertaining cancer outcomes in cohort studies [9].

Overall, the roles that cancer registries play in the control of cancer are critical. Registries foresee problems through continuously assessing cancer burdens, enabling the investigation of related risk factors, assessing the planning of appropriate intervention programs, and, finally, evaluating the outcomes of the intervention. Cancer registry data can also be used in epidemiological investigations, including descriptive studies to generate hypotheses and analytic studies to test the relationship between cancers and their contributing risk factors.

C. Data Quality in Cancer Registries

Data quality in registries must be monitored routinely through continuous surveillance activities to maintain accuracy, completeness, and timeliness. High quality data in population-based cancer registries are essential for the planning and evaluation of preventive and therapeutic services, as well as for the conduct of valid research. Better

health care planning and resource allocation can be achieved through the use of high quality registry data. Poor quality data have negative implications for health care practice, planning, and resource allocation [4]. If registry data are of poor quality, then the conclusions drawn likely will be invalid. As a result, those individuals planning activities for cancer control, including prevention and treatment programs, health care resource allocation, and overall public health in the community will be misled. In addition, consistency of data quality across cancer registries is essential for conducting comparative studies on the effectiveness, efficiency and cost of cancer control for health policy makers, health planners, and health care providers.

Data validity is an essential component in the quality control of registry data. It is defined as the proportion of cases in a registry with a given characteristic (e.g., cancer site, or age) which truly has the attribute [10]. The lack of accuracy in cancer registration refers to the situation when the data items collected are not necessarily correct. This can arise from the process of abstraction, transcription, coding, and even data entry errors in cancer registration [3]. The problem with timeliness relates to the time needed before diagnosed cases are registered on the registry database. It is expected that there will be negative impacts if registry data are not reported in a timely fashion on producing reports or on epidemiological studies.

The completeness of registration data may be defined as the extent to which all the incident cancers occurring in a target population are included in the registry. Ideally, completeness should be close to 100% so that the comparison of rates between registries reflects true differences in the risk of cancer and not artifacts of the registration process [10]. While it is important to achieve this goal, it is also important to avoid the

duplication of information on patients, and so most registries have sophisticated techniques to avoid such duplication [3]. If the degree of completeness of registration is low, comparisons among registry data cannot be interpreted as being attributable to different cancer risks in the various populations, and the trends observed in the incidence and mortality rates may not be used reliably in health planning. Therefore, completeness should be constantly monitored as part of the quality control procedures. The estimation of completeness is the main feature of this thesis and will be discussed in more detail.

D. Methods of Estimating Registration Completeness

The methods used to estimate completeness vary across registries because of the various case-finding mechanisms and registration procedures. The methods used for evaluating registration completeness can be grouped as suggested by the International Agency for Research on Cancer (IARC) [10-11]. A full assessment of the analysis should be on a site-specific basis even though each method can estimate the overall degree of completeness. Owing to various results obtained as a result of different registration mechanisms among registries, consideration should be given so that methods complement one another [10]. The main methods are described below:

1. ***Data Sources***

This group of methods involves the use of existing information in the registry database. Since completeness estimates using existing data sources in individual

registries are not absolute, the relative measures should be derived based on the comparison made with other registries.

1.1 Number of Sources/Notifications Per Case

Most population-based cancer registries use several case-finding procedures as a routine practice. These sources include reviews of pathology and autopsy reports, hospital patient-disease information systems, radiotherapy notes and death certificates. It is expected that greater completeness can be achieved with higher average numbers of sources per case in the registry database [10]. It is understandable that the possibility of registering cases would increase if there were more notifications per case reported to the registry.

The indicators include a comparison of the number of data sources per tumour or the number of notifications per tumour [12-14]. The estimates derived here are relative rather than absolute. Since the ACR uses pathology reports as a major source, and death certificates as a supplemental source, this method would not be applicable to the ACR.

1.2 The Death Certificate Method

The percentage of cases identified by death certificates only can be used as an indication of registration completeness [15]. The death certificate is one of the information sources most registries use to register cancer cases. Registries use death certificate notification (DCN) from Vital Statistics to match cases against their files and perform active follow-up on those cases when information is available only from death certificates. If there is no other information available but a death certificate, after inquires

have been made with physicians or hospitals, they are registered as DCO cases. However, policies on following up on DCN cases vary among registries.

Death certificates serve to detect cancer cases that were unregistered during the life of an individual. The indicator of completeness using DCN% is the proportion of cases first reported to the registry by sources other than the death certificate. It is therefore not sensitive for detecting cancers with low case fatality rates. Some types of cancer can have lower registration rates than others because some patients receive different types of medical care (e.g., melanoma). Thus, DCN% should be used on a site-specific basis. Finally, the indicator of DCN% may be affected by various policies on recording suspect cancer cases among cancer registries [10]. Although DCN% is better than the DCO% as a completeness indicator [10], we used the DCO% in this study.

1.3 The Percent Histological Verification of Diagnosis (HV%) Method

The completeness of registration also can be estimated through the percentage of cases being histologically confirmed [16-18]. This measure may overestimate completeness for registries that use pathology reports as the main means of case finding such as is used in the ACR, because cases diagnosed by other methods may be missed. A very high HV% therefore gives rise to a suspicion of under-registration (incompleteness). However, there is an increasing problem as imaging becomes more common in the diagnosis of cancer.

2. *Independent Case Ascertainment*

The methods described in this category take advantage of the availability of information from several different sources for the same cancer case to estimate registration completeness.

2.1 Comparison with Independent Data Sets

This method involves comparing cancer registry data to another comprehensive data source that includes information on diagnosis in which the data are relatively complete and are independent of registry data to identify potential unregistered cancer cases [5, 11, 19-39]. It is thought that this is the most straightforward and perhaps the most definitive method of determining completeness [40]. Case follow-back should be performed to confirm the diagnosis for unmatched cases. Results from the case follow-back should be incorporated into the results from the record linkage for completeness estimates. After incorporating the results from the case follow-back, the completeness of registration is estimated by dividing the number of records for which agreement has been achieved in the independent data set compared to the number of records in the registry, over a defined time period. Results from the comparison between two data sets should be analyzed and interpreted on a site-specific basis because of the various degrees of completeness reached among different sites.

2.1.1 Record Linkage

With advances in computer technology, it is feasible to compare registry data to another data set through computerized record linkage. Record linkage is a process of

matching records between two files. Records representing identical individuals can be linked, while others remain unlinked. The reason that record linkage is widely used nowadays is for its ability to connect different parts of the life experience of the same individual together. Doing so facilitates the conduct of epidemiological research such as in cohort and case-control studies. Thus, the method of computerized record linkage between registry data and another independent database is commonly used for this comparison [15, 40-41].

Record linkage has been used in Canadian health research mainly because of the existence of large administrative databases that support government-run, population-based health care insurance plans. It is useful for researchers to link these data files for epidemiological studies because these databases contain potential exposure and outcome information. The main advantage of record linkage is its cost-effectiveness [42].

In general, there are two approaches used in doing record linkage -- the deterministic and the probabilistic approaches. The major criterion concerning the adoption of either of these two approaches is the existence of unique identifiers between two data files. Deterministic linkage connects the records on the same individuals in two data files, based on the co-existence of unique identifiers. It provides a simple assessment of whether the two records are a true match based on the identifier linked. Unlinked records may be further matched by other personal identifiers such as name, sex, and date of birth to minimize the possibility of coding errors for unique identifiers in two data files.

The probabilistic approach uses numeric assessment to generate weights to classify whether the records in two data files represent identical individuals [43]. The cutoff

threshold weights can be determined, including one high and one low, based on acceptable probabilities for false matches and false non-matches. Those matches with weights higher than the high threshold weight are then considered matched and those matches with weights lower than the low threshold weight are considered non-matches. Matches with weights between two thresholds are deemed to be in a gray area and need clerical review [44-45].

Each method has advantages. Deterministic linkage is relatively simple and is used primarily when data are known to be complete and to have low levels of coding errors in the identifying information. Since probabilistic linkage involves generating weights (probabilities) for each potential link, this method is most advantageous when few variables are linked, data are incomplete, or coding errors are common [46]. However, probabilistic linkage is much more complex than deterministic linkage, and so accuracy versus simplicity becomes a trade-off when deciding on which approach to use [46-48].

The supplementary use of SOUNDEX codes to match records that belong to the same individuals in record linkage is popular [49-56]. The SOUNDEX program allows for the identification of names that are phonetically similar and thus avoids errors [57]. The reason for using of SOUNDEX codes is errors in names.

2.2 Re-screening of Cases

Re-screening of cases involves the procedure of reviewing case records for a certain period of time at one of the case-reporting facilities – usually a hospital - and comparing these results with those in a registry database. The degree of incompleteness is determined by the percentage of cases missed in the registry's routine procedures [10]. It

is assumed that fewer cases would be missed when the method of re-screening cases is used to identify cases in one of the routine case-reporting sources [58-61].

2.3 The Capture/Recapture Methods

The use of the capture-recapture method takes advantage of the multiple source reporting systems in most registries [62-67]. This method is used widely in wildlife population censuses [68]. The total number of animals in the population can be estimated through the procedure of counting marked animals recaptured after they were captured, marked, and released [10]. Registries with multiple data sources could adapt this method to evaluate the completeness of registration [63, 65, 69-70]. The multi-source reporting system in Ontario which includes hospital discharge summaries, pathology reports, Regional Cancer Centre Records, and death certificates makes the use of this method feasible for estimating the completeness of the Ontario Cancer Registry (OCR) [63].

2.4 The Incidence:Mortality (I:M) Ratio

The I:M ratio is a comparison of the number of newly diagnosed cases with a specific cancer to the number of deaths attributed to the same cancer in the same time period [71]. The I:M ratio should be stable over time if the cause of death is accurate and incidence and survival are in a constant state. For cancer sites such as the female breast, colon, skin, cervix, and testis, the I:M ratio will be larger than 1.0 because of good survival after diagnosis. An I:M ratio close to 1.0 will be found among cancer sites in the lung, bone, brain, liver, esophagus, and pancreas, having shorter survival. An I:M ratio less than what is expected can indicate under-registration. Thus, we can estimate registry

completeness through an examination of the I:M ratio for cancer sites with known average survival rates [10].

The North American Association of Central Cancer Registries (NAACCR) uses the I:M ratio as an independent case ascertainment method to estimate the completeness of its member registries across North America. The method for obtaining registry completeness estimates in the publication "*Cancer in North America*" (*CINA*) by the NAACCR is used to calculate the expected incidence rate for individual registries based on the Surveillance, Epidemiology, and End Results Program (SEER) incidence rate and the National Mortality Data Base (NMDB) mortality rate in the U.S. It is used to derive completeness estimates by comparing the observed incidence rate to the expected incidence rate for individual registries [72]. The interpretation and analysis of the results using this method should be on a site-specific basis. The ACR has been honoured with the Gold Standard by NAACCR with the use of this method for its data in the years of 1995, 1996, and 1997.

The method of I:M ratio used to estimate the completeness in NAACCR is as follows [72]:

$$\text{Expected } I_{s(ij)} = (M_{s(ij)})(I_{SEER(ij)} / M_{U.S.(ij)})$$

$$\text{Expected } T_S = \sum_{i=1}^{2} \sum_{j=1}^{N} \text{Expected } I_{S(ij)}$$

- I = age-adjusted incidence rate in gender i and site j category
- M = age-adjusted mortality rate in gender i and site j category
- s = State, SEER area, province or territory
- SEER = combined nine historical areas in the SEER program
- $U.S.$ = United States
- T_s = age-adjusted incidence rate for total sites with both gender combined

Cancers of the breast and prostate are not included for the estimate of T for total sites owing to the lack of stability in the I:M ratios caused by the dramatic increase in screening for these diseases in some regions.

The percent completeness is calculated by dividing the observed (or reported) age-adjusted incidence rates for all the sites for both genders (Observed T_s) by the expected age-adjusted incidence rate for all the sites for both genders (Expected T_s).

$$\% \text{ Completeness} = \text{Observed } T_s / \text{Expected } T_s$$

The number of duplicate records in the data set is calculated using the NAACCR estimate of duplicates, based on the registry's results after completing the protocol for *Assessing Duplicate Cases* [72]. The NAACCR completeness estimate is further adjusted for duplicates using the following equations:

Adjusted % completeness = Observed $T_s - D_s$ / Expected T_s

- Observed T_s = age-adjusted incidence rate for total sites
- Expected T_s = estimated age-adjusted incidence rate for total sites if completeness is 100%
- D_s = age-adjusted incidence rate of duplicate records for total sites

For registries that did not complete the *Protocol for Assessing Duplicate Cases,* the NAACCR adjusted estimate for completeness is omitted from the registry description.

The fact that the information on death certificates from vital statistics is less precise and accurate than that recorded in most of the registries should be recognized. In addition, changes in incidence and mortality rates as a result of new screening and treatment programs can have impacts on the I:M ratio. Thus, care should be taken in interpreting results when the I:M ratio is used to estimate registry completeness.

3. ***The Historic Data Method***

3.1 Incidence Rates over Time

This method compares the incidence rates with those from the same registry in earlier periods. Results are further standardized to appropriate reference groups and are compared on a site-specific basis by sex, age group, and other demographic factors. Changes in incidence rates over time and changes in underlying incidence may reflect new screening programs or practices in cancer registration [73-76].

3.2 Comparison of Incidence in Different Populations

Comparison of incidence in different populations covered within the registry can reflect either true differences in cancer risks or variations of completeness in cancer registration. Similar incidence rates would be expected when comparing similar populations within the registry or when comparing registries with similar demographic populations [10]. It is noted that the data comparability among registries should always be taken into account when making such comparisons.

3.3 Age-specific Incidence Curves

Observation of the pattern of age-specific incidence curves can provide an indication of registration completeness. Changes in these curves comprising data from different birth cohorts can reflect changes in risk or the indication of possible changes in registration practice [10]. Such cohort effects can be detected only when data are available from several time periods. Therefore, the shapes of age-specific incidence curves are important indicators of possible under-ascertainment [77-78].

4. ***A Statistical Modeling Technique***

One method based on the logical flow of data in the registration system has been proposed to assess registry completeness. It takes advantage of using existing data in the registries by incorporating the probabilities of survival $s(t)$, the registration of the cancer during the patient's life $1-u(t)$, and of the mention of cancer on the death certificate of the cancer patient who dies $m(t)$. Completeness of registration is estimated based on the three functions mentioned above [79].

Two criteria need to be satisfied before this method can be used routinely in registries: death certificates specifying cancer as a cause of death must be received for all patients who have deceased, and patients must not be registered from sources other than death certificates after death. The method proposed seems applicable to most registries that fulfill these two assumptions. However, the unavailability of information on exact date of registration for patients in the ACR makes the estimate of the probability $u(t)$ impossible. Thus, this method is not applicable to the ACR. Although this method has not been adopted and evaluated in other registries or in major registry associations such as International Association of Cancer Registries (IACR) or SEER, it is ideal to incorporate the results from this method with results from the methods chiefly used.

E. Cancer Registries in Canada

Cancer registries have been in operation in all Provinces in Canada for over 25 years, and are now operational in all the Provinces and Territories. These registries use different methods to collect cancer data but try to follow the same rules of coding suggested by the Canadian Cancer Registry (CCR). While some are based only on electronic linkages, others use manual abstraction or a combination of the two [80]. Information on cancer incidence and mortality is sent from the provincial and territorial cancer registries and offices of vital statistics, which send their data to Statistics Canada for compilation at the national level to form the CCR data base [81].

Cancer registration procedures vary from province to province across Canada. This variation in procedures may lead to variations in the completeness, timeliness, and

accuracy of cancer registration. The extent of these differences is probably greater than is commonly recognized [82]. Uncertainty about data quality introduces a difficulty in interpreting regional differences, temporal changes in cancer rates, and outcomes. Most quality management activities in Canadian Cancer Registries currently are passive and done on an *ad hoc* basis. Alberta, Ontario, and Nova Scotia have done some systematic data audits over the last few years to look at the accuracy of the core data elements and the completeness of case ascertainment in cancer registration [80].

The CCOCS has developed a National Action Plan that describes the issues that need to be addressed, and details the priority actions required to put a prototype cancer surveillance system into place. The need to conduct national studies to determine the completeness and accuracy of cancer registry databases is recognized and prioritized by the Quality Management Working Group in CCOCS. In the initial phases, federal and national funding would be required to establish the infrastructure required for the development and building of an enhanced cancer surveillance system, with the provinces being active participants from the inception and eventually taking on stronger roles as time progresses [83].

A collaborative pilot study funded by Statistics Canada is currently being undertaken to determine the most appropriate, consistent and cost-effective system to evaluate case ascertainment for provincial and national cancer registries. Three pilot sites have been chosen, the provinces of Ontario, Manitoba, and Alberta, because of the different case-finding procedures, various registration processes, and different population structures they have. Results of this study in Alberta will contribute to the national project.

F. The Alberta Cancer Registry (ACR)

The ACR is legislated to capture information on every incident case of cancer diagnosed in Alberta. The ACR is a computer file of records on all patients who have been diagnosed in Alberta with a malignant tumour (in situ as well as invasive) or with a benign or borderline nervous system tumour (e.g., meningioma) or a borderline hematologic disorder. The International Classification of Diseases for Oncology (ICD-O) 2nd Edition is used by the ACR to record malignant information for cancer patients. The registry contains identifying information, details of diagnosis, treatment of the tumour, and of death. The information entered into the registry is abstracted from various documents which are maintained in the medical record, such as the discharge summary, progress notes, pathology reports and surgery reports [57]. Although ACR also receives information in a non-routine fashion from hospital discharge summaries and other sources, pathology reports and information through Vital Statistics are two main sources of identifying and registering incident cancer cases in Alberta. According to the Cancer Registry Regulation of the Cancer Programs Act in Alberta, all physicians and laboratories must provide information about cancer cases diagnosed in Alberta to the Alberta Cancer Board (ACB). This information would include patient, tumour, and tumour-specific data.

Information from Vital Statistics on all deaths occurring in Alberta is routinely linked to the ACR to provide information on cancer cases that have died, and to identify those potential cancer cases that may not have been captured through pathology reports.

G. The Effect of Different Coding Rules on Estimates of Registry Completeness

Registries use different mechanisms to register cases, along with various coding practices and rules. Coding practices using different rules would definitely have an impact on the comparability of cancer registry data. That is especially true for comparing incidence and registration completeness among registries. There are two commonly used rules for coding malignant neoplasms; the one is suggested by the SEER, and the other is suggested by the IARC and the IACR. The ICD-O2 is the disease classification scheme that these two programs are based on. They are similar in most respects with the exception of the rules for coding multiple tumours. For both SEER and IARC coding rules, the first 3 digits of the ICD-O2 topography designate different sites, except for those topography codes listed in Table 2.1, and the exceptions listed below [84]:

- the topography of cancers of the colon (C18), anus and anal canal (C21), bones, joints and articular cartilage (C40, C41), melanoma of the skin (C44,M872-9), peripheral nerves/autonomic nervous system (C47) and connective, subcutaneous and other soft tissues (C49). In SEER rules, a difference in the fourth character of the topography code for these cancers would be designated a different site. Under IARC coding, rectosigmoid junction (C19) and rectum (C20) are considered one site, as are the bones, joints and articular cartilage (C40, C41).
- the morphology of the cancers. In SEER rules, the first three digits of the morphological code designate the different histological types, except for lymphatic and hematopoietic diseases. IARC defines 10 groups to be histologically different (Table 2.2).

- the timing of the tumour. SEER would code as different tumours any tumours occurring more than 2 months apart that the clinician has not designated as a recurrence. IARC has no time constraint.

Table 2.1 ICD-0-2 Topology Codes to be Considered One Primary Site When Determining Multiple Primaries for both ACR & IARC Coding

ICD-0-2 Codes	Site Groupings
C01 C02	Base of tongue Other and unspecified parts of tongue
C05 C06	Palate Other and unspecified parts of mouth
C07 C08	Parotid gland Other and unspecified major salivary glands
C09 C10	Tonsil Oropharynx
C12 C13	Pyriform sinus Hypopharynx
C23 C24	Gallbladder Other and unspecified parts of biliary tract
C30 C31	Nasal cavity and middle ear Accessory sinuses
C33 C34	Trachea Bronchus and lung
C37 C38.0 C38.1-.3 C38.8	Thymus Heart Mediastinum Overlapping lesion of heart, mediastinum, and pleura
C38.4	Pleura
C51 C52 C57.7 C57.8-.9	Vulva Vagina Other specified female genital organs Unspecified female genital organs
C56 C57.0 C57.1 C57.2 C57.3 C57.4	Ovary Fallopian tube Broad ligament Round ligament Parametrium Uterine adnexa
C60 C63	Penis Other and unspecified male genital organs
C64 C65 C66 C68	Kidney Renal pelvis Ureter Other and unspecified urinary organs
C74 C75	Adrenal gland Other endocrine glands and related structures

Table 2.2 Groups of malignant tumours considered to be histologically different for the purpose of defining multiple tumours using IARC rules

GROUP	CARCINOMAS	ICD-0 MORPHOLOGY CODE
1	Epidermoid carcinomas	805-813
2	Adenocarcinomas	814, 816, 818-822, 825-850, 852-855, 857, 894
3	Other specified carcinomas	803-804, 815, 817, 823, 824, 851, 856, 858-867
4	Unspecified (Carcinomas NOS)	801-802
5	Sarcomas and other soft tissue	868-871, 880-892, 899, 904, 912-913, 915-934, 937, 954-958
6	Other specified types of cancer	872-879, 893, 895-898, 900-903, 905-911, 935-936, 938-953, 972-974, (976 for ICD-O-2 only)
7	Lymphomas	959-971 (975 for ICD-O-1 only)
8	Leukaemia	980-994
9	Kaposi's sarcoma	914
10	Unspecified types of cancer	800 (999 for ICD-O-1 only)

The IARC/IACR rules are generally used to submit data for *Cancer Incidence in Five Continents*, whereas SEER rules are generally used to define cancer sites for producing *CINA,* published by the NAACCR. SEER program participants use SEER rules to submit data to the National Cancer Institute (NCI) in the United States. Most of the cancer registries in the North America adopt these coding rules, including the ACR (except for the date of diagnosis), which makes cancer registry data consistent and comparable across North America. The CCR also uses SEER coding rules except for definition of the date of diagnosis and the timing of tumours occurring 3 months apart being considered as a new primary in the CCR as opposed to 2 months apart in the SEER. However, it is possible to apply IARC coding rules to the SEER data to calculate the number of cases that would have been registered in SEER registries using IARC rules.

With increasing survival after treatment for several forms of cancer, and the use of chemotherapeutic agents which are themselves carcinogenic in the treatment of malignant disease [85-87], it is estimated that at present some 5% of all cancer patients develop a further independent primary cancer [88-89]. Thus, the importance of coding practice using different rules on multiple tumours should always be taken into consideration while making data comparisons among registries, including estimates of completeness.

H. Confidentiality and Ethics

There is concern about protecting individual privacy. A dilemma emerges between the need to protect individual privacy and the public good that flows from registry data. It has been especially obvious in recent years because computer technology and

information exchange in research is much advanced. Human dignity, informed consent, and privacy and confidentiality of individuals are to be respected. Yet a balance needs to be struck between minimizing harms to individuals and maximizing benefits to the whole community based on moral considerations in research ethics [90]. In practice, the public good can be maximized through the use of personal health information for conducting research; in the meanwhile, the potential harms to individual privacy from the disclosure of personal health information should be minimized. The ethical issue of confidentiality in conducting research on humans should always be prioritized [90].

The introduction of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* to replace guidelines from the Medical Research Council (MRC), the Natural Sciences and Engineering Research Council (NSERC), and the Social Sciences and Humanities Research Council (SSHRC) certainly has profound impacts on research. It has become a joint guideline for ethical consideration for conducting health research in this country [91].

The conduct of this study involved the use of identifying information both in ACR and AHW data files for record linkage, including the data fields of personal health number (PHN), surname, initial, sex, date of birth, and place of residence. Although studies with the use of cancer registry data do not interact with participants directly as in community-based studies, ensuring the ethical considerations are met is no less important. It is the responsibility of the registry to ensure that the confidentiality of the registry records will be preserved, whether the record linkage is implemented inside or outside the registry [3]. The following are ethical issues related to research using

population-based registry data, including the principles of respect for autonomy, beneficence, non-maleficence, and justice [92].

Respect for autonomy involves respect for the rights of individuals and groups to make decisions for themselves. In public health, it is accepted that there are some situations in which autonomy must be limited to protect the public's health, such as the requirement of immunization against communicable diseases [90]. But in most instances, autonomy is a central ethical principle. In this study, for example, information on all cancers diagnosed in Alberta should have been provided to the ACR under the regulation of the Alberta Cancer Programs Act. The goal of improving and the better control of cancer, for agencies to provide data through cancer research is a reason to respect autonomy to protect the public's health.

Conflicts can always arise in research when individuals' rights to privacy are breached with the goal of obtaining important knowledge to improve public health. In studies using registry data, it is the researcher's responsibility to ensure the privacy and confidentiality of participants by using aggregate data rather than individual data as much as possible, or using the individual data with identifying information removed [93]. Whenever possible, the use of data encryption is suggested to store or to transmit data to minimize the risk of disclosure. However, all possible efforts should be made to minimize the chances of identifying individuals by linking the identifying information on different data sets by unauthorized persons.

In addition, investigators should obtain approvals from the registry, related agencies, and the appropriate ethics review committee before commencing studies.

Investigators should sign a confidentiality statement to prevent unauthorized release of such information. Data should be stored in computers with password protection, and should be accessible only to research-related personnel. Any paper copy should be secured in locked offices. These procedures more assure respect for the autonomy of individuals and also of the partner institutions.

Beneficence & non-maleficence. The principle of non-maleficence requires that researchers minimize risk to participants; the principle of beneficence requires that researchers maximize the potential benefits both to cancer patients and to society. In this study, since cancer registry data is a primary resource for planning and evaluating health services for the prevention, diagnosis and treatment of cancers, data quality in the ACR could be improved through the findings from this study. Thus, it is beneficial to society and does not harm or adversely affect cancer patients.

Justice. The ethical principle of justice concerns the fair distribution of benefits and burdens of research among potential patients. Based on the utilitarian theory of ethics, the registry functions to provide the greatest good for the greatest number of people. The public health benefit should be maximized with the requirement that all citizens receive an equal share of the benefits from the research. In this ACR case ascertainment study, the benefits of data quality improvement to facilitate cancer control should be equally distributed among the population. It is expected that an equal share of benefits, like better health care resource allocation, will be achieved based on research that will more assure high quality data in the ACR, which is the objective of this study.

I. Legislation on Accessing Cancer Information

To reduce the cancer burden in Canada, a comprehensive cancer surveillance system is essential. The system serves to detect changes in cancer risks in populations, to manage the health care system, to evaluate prevention and treatment programs, and to identify targets for the further control of cancer [81]. Appropriate legislation on accessing health information facilitates the establishment and management of a surveillance system. Such legislation involves data collection procedures, appropriate information exchange and access to health data. A balance needs to be maintained between the potential harm resulting from breaching individual privacy and the public good achieved through the use of health data.

Currently, the right to privacy is upheld by federal and provincial legislation and by broad sets of guiding principles. The scope of provincial legislation differs from province to province. For example, in Quebec, legislation covers personal information held in both the public and private sectors. However, other provinces extend coverage only to the public sector. In the 1980s and early 1990s, six of Canada's provinces and the Northwest Territories passed comprehensive measures in the form of the *Freedom of Information and Protection of Privacy* (FOIPP) legislation. In addition to this general legislation, individual agencies such as Statistics Canada have prepared detailed guidelines for protecting personal information [94].

The enforcement of FOIPP legislation has had a significant impact on research in Alberta. It includes the regulations concerning accessing personal information by researchers and interested parties and protecting it from inappropriate disclosure. Health information is of particular concern. Access to and disclosure of personal health

information by health practitioners, health facilities, health records, and researchers are regulated in detail by FOIPP. For example, the fact that FOIPP restricts cancer information exchange among provinces impedes access to information on out-of-province residents for follow-up. The *Health Information Act* in Alberta specifically mandates that personal health information may be provided to researchers and interested parties by health practitioners and health facilities only under supervision [95]. The protection of the privacy and confidentiality of such information is warranted. The use and disclosure of cancer information have been defined in the *Alberta Cancer Programs Act* in detail [96].

Recommendations about legislation on cancer information have been made by the Legislative Working Group in the CCOCS: "It will be necessary to undertake a review of the formal privacy, access and security policies across the cancer registries/agencies and health care providers within Canada." The group has recommended focusing first on cancer registries, followed by the cancer agencies, and finally other defined health care providers. According to CCOCS, "It will be determined whether such policies are consistent with provincial/territorial/federal privacy legislation and other privacy, access and security standards. It is hoped that this review will lead to the development of a model policy concerning privacy, access and security for both primary and secondary uses of data." [97].

J. Methods Appropriate to Estimate ACR Completeness

Based on the situation in Alberta and the methods for determining completeness estimates discussed above, it is appropriate to compare the ACR data to an independent data set to estimate ACR completeness. I:M ratios and indicators of HV% and DCO% derived from ACR data can be used to provide indicators of ACR completeness. In addition, case re-finding in a hospital can be conducted to supplement these methods.

1. *Comparisons between AHW and ACR Data*

The method for this study will be to compare ACR data to hospital discharge and day procedure data maintained by AHW. Records in AHW data for the years of 1994-96 will be matched to ACR data and will be checked to determine whether cancer diagnoses in AHW data reflect those recorded in ACR data. Owing to the co-existence of a unique identifier -- the PHN -- in both data files, as well as evidence of fairly complete data with fewer coding errors in the ACR [98], it is appropriate to adopt a deterministic approach for this record linkage. Completeness will be estimated based on the percentage of AHW records that link to the ACR with agreed cancer diagnoses.

2. *Use of Incidence:Mortality (I:M) Ratios*

I:M ratios provide an absolute measure of registry completeness. They can be used to estimate ACR completeness based on the method used in the NAACCR publication – *CINA* [72]. The use of SEER incidence and U.S. mortality data as a standard reference may not be appropriate for Canadian registries because of differences in coding rules for coding multiple primaries and in the survival experience between the two countries [84].

3. *Use of Indicators of Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%)*

HV% and DCO% are relative measures of registry completeness. They are derived from ACR data and are compared to a reference group that has a similar mechanism for registering cases. They also can be compared to ACR data derived from previous years. The differences and changes in these indicators provide information on ACR completeness. Because some studies have shown unreliability in the use of these indicators [70, 99], results obtained using these indicators will be considered in conjunction with the results from the method of independent case ascertainment.

4. *Case Re-finding in the University of Alberta Hospitals (UAH)*

Ideally, the case re-finding exercise in a hospital can be used to supplement the methods identified above. The UAH was chosen for this exercise. Consideration was given to the ease of access and the costs associated with this exercise at the UAH. Owing to the centralization and computerization of health information in the Capital Health Authority (CHA), outpatient records of the UAH for the years 1994-96 indicating cancer diagnosis can be provided by the CHA.

A deterministic record linkage will be conducted to match cases from the UAH outpatient data to cases in the ACR. The rationale and criteria used for this record linkage will be similar to those in the AHW-ACR linkage. The completeness estimate will be derived in a similar way to that for the AHW-ACR linkage.

With the various methods proposed above, the final estimate of ACR completeness will be reconciled based on the strengths and weaknesses of each method respectively. Comparisons among methods will be made. Advantages and disadvantages among methods used will be further discussed, resulting in a final set of recommendations.

CHAPTER III. MATERIALS

The data files used in this study include data sets from the ACR, hospital discharge and day procedure records from AHW, and outpatient records from the UAH provided by the CHA. For record linkage between AHW and the ACR, records with any cancer diagnosis (ICD-O2, C00-C80) for all years in the ACR, as well as records with any cancer diagnosis (ICD-9CM, “140-208”) for the years of 1994-96 from AHW were used. Consideration was given to the possibility that cancer patients in AHW data were registered previously in the ACR. Incidence and mortality data for the years 1994-96 in the ACR were used to derive I:M ratios, together with SEER incidence and NMDB mortality including only white populations in the U.S. as the standard registry for completeness estimates. The 1970 U.S. population was used as a reference group for age-standardized rates. For completeness indicators such as HV% and DCO%, records in the ACR were used. The same data set in the ACR for the AHW-ACR record linkage was used to match outpatient records with cancer diagnoses for the years 1994-96 in the UAH database in a case re-finding exercise.

A. The Alberta Cancer Registry

There are three major types of information collected by the ACR: patients’ identifying information (Registration Segment), information about malignancy (Malignancy Segment), and information on deceased cases (Death Segment).

Registration Segment

This segment contains information that enables us to identify the patient. It is important to avoid creating more than one record for the same patient. Therefore, it is crucial that the registration segment contains correct information. These variables used in this study are described below.

Data Fields	Definition	Values
ACB_NO	The number which uniquely identifies each case registered on the ACR	
LST_NAME	The surname of the patient	
FST_NAME (MID_NAME)	The first (middle) name of patient as given at birth	
NAME_TYP	The type of name which has been recorded	Current, preferred, birth, and other
DOB	The date the patient was born	
SEX	The gender of the patient	
PHN	The Personal Health Number of the patient	
NUM_MAL	The total number of malignant tumours the patient has	

Malignancy Segment

The variables in this segment must be entered for every new primary malignancy. One single patient may have more than one malignancy segment present in the database. Each malignancy is indicated by the variable "malignancy number."

Data Fields	Definition	Values
ACB_NO	The number which uniquely identifies each case registered on the ACR	
MAL_NO	A counter indicating the sequence number of the malignancy	
ICDO_TOP	The code representing the anatomical localization of the malignancy (The ICD-O-2 is used)	
ICDO_MOR	The code representing the morphology that makes up the major/specific component of the malignancy (The ICD-O-2 is used)	
DIAG_MET	The most definite diagnostic procedure by which a malignancy is diagnosed within 2 months of the earliest known encounter with the health care system for that tumour	Histology, cytology, laboratory, surgery, radiology, clinical, death certificate, and unknown
HIST_CNF	Has this cancer been microscopically confirmed ?	Yes, no, and unknown
DIAG_DTE	The date of diagnosis represents the date of the most definite diagnosis made	
PC_DIAG	The postal code of the residence of the patient at the time of diagnosis	
SGC_DIAG	Standardized geographic codes for places of residence when cancer cases are diagnosed	

Death Segment

Capturing deaths in the cancer registry allows for estimating mortality, survival, and prevalence. Also, the death certificate sometimes mentions cancer in a patient as yet unknown to the cancer registry; i.e., death certificates are used in case finding. The proportion of cancers diagnosed by DCO is a measure of the thoroughness of a cancer registry's follow-back on DCN cases. Information from Vital Statistics is compared to that recorded in the ACR. Medical charts are reviewed to determine the underlying cause of death.

Data Fields	**Definition**	**Values**
ACB_NO	The number which uniquely identifies each case registered in the ACR	
DTH_DTE	The date of death of the patient as mentioned on the death certificate (or on the Vital Statistics data file)	
DTH_PC	The postal code of the residence of the patient at the time of death	
DTH_CS1	The underlying cause of death is the disease or injury which initiated the chain of events leading directly to the death of the patient (or the circumstances of the accident or violence which produced the total injury) (ICD-9 is used). The way to determine the underlying cause of death in the ACR is to compare the information from Vital Statistics to that recorded in the ACR. Medical charts are reviewed to make a judgement whenever disagreement exists.	

1. **The AHW-ACR Record Linkage and UAH Case Re-finding**

Cases with cancer diagnosis as determined by topography and morphology (ICD-O2, C00-C80) for the years between 1994 and 1996 in the ACR were used in the study. Because consideration was given to the possibility that cancer patients in AHW data were registered previously in the ACR, data from the ACR for the years prior to the end of 1996 were used in the linkage. This comparison, however, may result in multiple primary cancers that develop in the same site being missed. Only Alberta residents were included in the study. Although all sites were investigated, cancers of the female breast, prostate, pancreas, melanoma of the skin, colon-rectum, lung, and leukemia were of particular interest.

2. **The Incidence:Mortality (I:M) ratio**

For the years 1994-96, cases with an invasive cancer diagnosis as determined by topography and morphology (ICD-O2) in the ACR were used to derive incidence. For mortality, cases that died of cancer as determined by the underlying cause of death for the years of 1994-96 were used. Only Alberta residents were included in the study for both incidence and mortality. Although all sites were investigated, cancers of female breast, prostate, pancreas, melanoma of the skin, colon-rectum, lung, and leukemia were of particular interest.

In the analysis to explore the discrepancies in completeness estimates using I:M ratio with different coding rules, data from the ACR, SEER cancer incidence, and the NMDB in the U.S. for the years 1994-96 were used. The cancer sites that were analyzed are those for which completeness is estimated in the *CINA* publication together with

female breast and prostate cancer. The ACR has used SEER rules since 1994. IARC coding rules were applied to Alberta data to calculate the number of cases that would be registered. The 1970 U.S. population was used as a reference group for age-standardized rates. SEER cancer incidence and the NMDB mortality including only white populations in the U.S. were used as the standard to calculate expected incidence for the ACR and to estimate registration completeness.

3. **Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%)**

For the years 1994-96, HV% and DCO% of cancer cases were derived based on the method of diagnosis recorded in the ACR. Cancer sites of the female breast, prostate, pancreas, melanoma of the skin, colon-rectum, lung, and leukemia were investigated.

B. Alberta Health and Wellness Data

Records of hospital discharge and day surgery files for the years 1994-96 from AHW, which contained a diagnosis of cancer (ICD-9CM, “140-208”), were provided to the ACR. Diagnostic codes of “V10” were used to code cases with personal histories of malignant neoplasm. However, because the practice in coding “V10” was not necessarily reliable in the AHW file, cancer cases diagnosed with “V10” codes were excluded. The corresponding records from the AHCIP registration files were also provided. Only Alberta residents were included in the study. The following are data fields in the AHW files used in the AHW-ACR linkage.

Day Procedure & Hospital Discharge Files

Data Fields	Definition
PHN	The Personal Health Number of the patient
SEX	The sex of the patient
DOB	The date the patient was born
PCODE	The postal code of the residence of the patient at the time of diagnosis
HOS_NO	The hospital where the patient was treated
DI_CD (1-16)	Diagnostic codes of the patient (The ICD-9-CM Diagnosis and Procedure Manuals are used)
PR_CD (1-10)	Procedure codes of the patient (The ICD-9-CM Diagnosis and Procedure Manuals are used)
DIS_STA	Discharge status of the patient (for day procedure)
DIS_DTE	The date the patient was discharged

Alberta Health Registry File

Data Fields	Definition
PHN	The Personal Health Number of the patient
LST_NAME	The surname of the patient
FST_NAME	The first name of the patient
SEX	The sex of the patient
DOB	The date the patient was born
PC_CODE	The postal code of the residence of the patient at the time of diagnosis

C. University of Alberta Hospitals Data

Service logs from the UAH with cancer diagnosis (ICD-9CM, “140-208”) for the years 1994-96 were provided by the CHA. These outpatient records include information on cancer diagnosis, as well as identifying information for cases. Only Alberta residents were included in the study. The following are data fields in the UAH outpatient files used in this exercise.

UAH Service Log

Data Fields	**Definition**
D_EVENT	The date the service was provided
PHN	The Personal Health Number of the patient
LST_NAME	The surname of the patient
FST_NAME	The first name of the patient
SEX	The sex of the patient
DOB	The date the patient was born
P_CODE	The postal code of the residence of the patient at the time of diagnosis
DI_CODE	Diagnostic codes of the patient (The ICD-9-CM Diagnosis and Procedure Manuals are used)

D. U.S. SEER Incidence and NMDB Mortality Data

1. **SEER Incidence**

Combined cancer incidence data including only white populations for the 9 historical SEER areas for the years 1994-96 were obtained from the SEER public use data CD-ROM produced by the NCI in the U.S., with the exclusion of the San Jose/Monterey and Los Angeles areas. The SEER Program currently collects and publishes cancer incidence and survival data from 11 population-based cancer registries and three supplemental registries covering approximately 14 percent of the U.S. population. Cancer sites used included those used in the *CINA* publication together with those of female breast and prostate.

2. **NMDB Mortality**

Cancer mortality data for U.S. registries including only white populations for the years 1994-96 were obtained from the National Center for Health Statistics of the Centers for Disease Control and Prevention (CDC), as provided to the NCI in the U.S. Cancer sites used included those used in the *CINA* publication together with those of female breast and prostate.

CHAPTER IV. METHODS

One of the methods used in this study was a record linkage between AHW and the ACR data sets. This linkage involved the comparison of the ACR data to an independent data set, which was the hospital discharge and day procedure file maintained by AHW for the years 1994-96. Completeness of the ACR was estimated based on the percentage of cases that the ACR failed to register, but which existed in the AHW data file. I:M ratios were also calculated to estimate the ACR completeness and to explore the discrepancies between completeness estimates resulting from the adoption of different coding rules among registries, based on the NAACCR method. Indicators of HV% and DCO% were derived from the ACR data in order to be compared to the results from the AHW-ACR linkage. In addition, a case re-finding exercise was conducted at the UAH to supplement the above methods. Although all sites were investigated, cancers of the female breast, prostate, pancreas, melanoma of the skin, colon-rectum, lung, and leukemia were of particular interest.

A. The AHW-ACR Record Linkage

This method compared the ACR data set with an independent data set from AHW. This data set included records from the hospital discharge and day procedure files with a discharge diagnosis of cancer for the years 1994-96. The rationale here was to check whether all the cancer cases diagnosed in Alberta were registered in the ACR. Procedures used in this record linkage were to check whether cases in the AHW data were registered

in the ACR data as a first step, and to further check whether the diagnostic information agreed if the case coexisted in both files. Cases that existed in the AHW files with the diagnosis of cancer but did not appear in the ACR files, as well as cases in the AHW files where the diagnostic information did not agree with that recorded in the ACR files, were followed back by ACR staff to confirm the diagnosis.

Special attention was focused on cases with multiple cancer diagnoses. Cases with more than one cancer diagnosis in the ACR had more than one record in the malignancy segment. The data field of "NUM_MAL" in the registration segment could also be used to identify these cases. For cases with more than one cancer diagnosis in the AHW data set, multiple records were created to facilitate record linkage. As long as every cancer diagnosis pertaining to a case in the AHW data set was reflected in the ACR data set, it was considered as a matched case, despite the possibility that not all the cancer diagnoses for a case in the ACR data set matched those in the AHW data set.

The following steps were used for the record linkage between AHW and the ACR data.

STEP 1

Link data sets between AHW and the ACR by the unique variable "PHN" coexisting in both files.

STEP 2

Check demographic information of date of birth and sex for those linked records in Step 1 to make sure those records are identical.

STEP 3

For those identical records in Step 2, check their diagnostic information to see whether they agree with that recorded in the ACR.

STEP 4

Create a data set for those unlinked records in AHW data in Step 1 and in Step 2, and link them to the Alberta Health Registry by "PHN" to get more identifying information concerning cases, in particular - last name and first name.

STEP 5

For those linked records in Step 4, link them back to the ACR data set by "SOUNDEX," "INITIAL," "SEX," and "DOB." Linked records were examined visually based on last name, first name, sex, and date of birth to make sure they were identical cases.

STEP 6

For those linked records in Step 5, check their diagnostic information to see whether they agreed with that recorded in the ACR

STEP 7

For those unlinked records in Step 4 and in Step 5, as well as those records with diagnostic information disagreeing with that recorded in the ACR in Step 3 and in Step 6, case follow-back was undertaken by the ACR staff. After going through available medical charts, cases resulting in questions concerning the diagnosis and those

unregistered cases were followed back by writing letters to the hospitals to confirm the diagnosis.

Differences in coding practice resulting from using ICD-9-CM (based on ICD-9) on AHW data and by using ICD-O 2nd Edition (based on ICD-10) on ACR data were taken into account in checking for agreement on diagnosis. A conversion table was used to check the diagnostic information in two data files. Doing so provided for the identification of cases with disagreement on diagnosis in two data files. Manual reviews were also conducted to resolve those cases with a questionable diagnosis.

Care was taken for cases with diagnostic information including ICD-9-CM "196", "197", "198", and "199.0" in the AHW data set in this linkage. Since these ICD-9-CM categories were used to code the metastatic malignant tumours, cases with these diagnostic codes were considered matched if they had any cancer diagnosis registered in the ACR. These cancer diagnoses in the AHW data set could have been based on cancer sites with unknown origins, resulting from metastasis from the primary tumours. Therefore, it was reasonable to exclude them from the list of unlinked records.
In addition, some of the diagnostic codes in the AHW data (ICD-9CM) and in the ACR data (ICD-O2) were considered equivalent (Table 4.1).

When the follow-back on unlinked cases is completed, cases that are really cancer cases but do not appear in the ACR data will be registered. Additional tumour information will be added to the ACR for cases whose diagnoses do not agree between the ACR and AHW data sets. After the results from the follow-back on unlinked cases are incorporated, the incompleteness of the ACR can be estimated by dividing the

number of unlinked records from Step 3 and Step 6 as well as unlinked records from Step 4 and Step 5 by the total number of records in the ACR for a certain period of time. ACR completeness can be estimated by levels of case coverage, which is defined as 1 – [unlinked AHW cancers/(number of ACR cancers + unlinked AHW cancers)], and obtained as 1 – incompleteness. Case coverage on a site-specific basis was also estimated accordingly. 95% confidence intervals were calculated using a binomial formula. Owing to the slow process of follow-back for those unlinked cases, results from the case follow-back can only be incorporated into the AHW-ACR record linkage results when it is complete. Thus, the case coverage has been reported in the results.

* Please refer to Figure 4.1 for this record linkage.

Figure 4.1 AHW-ACR Record Linkage

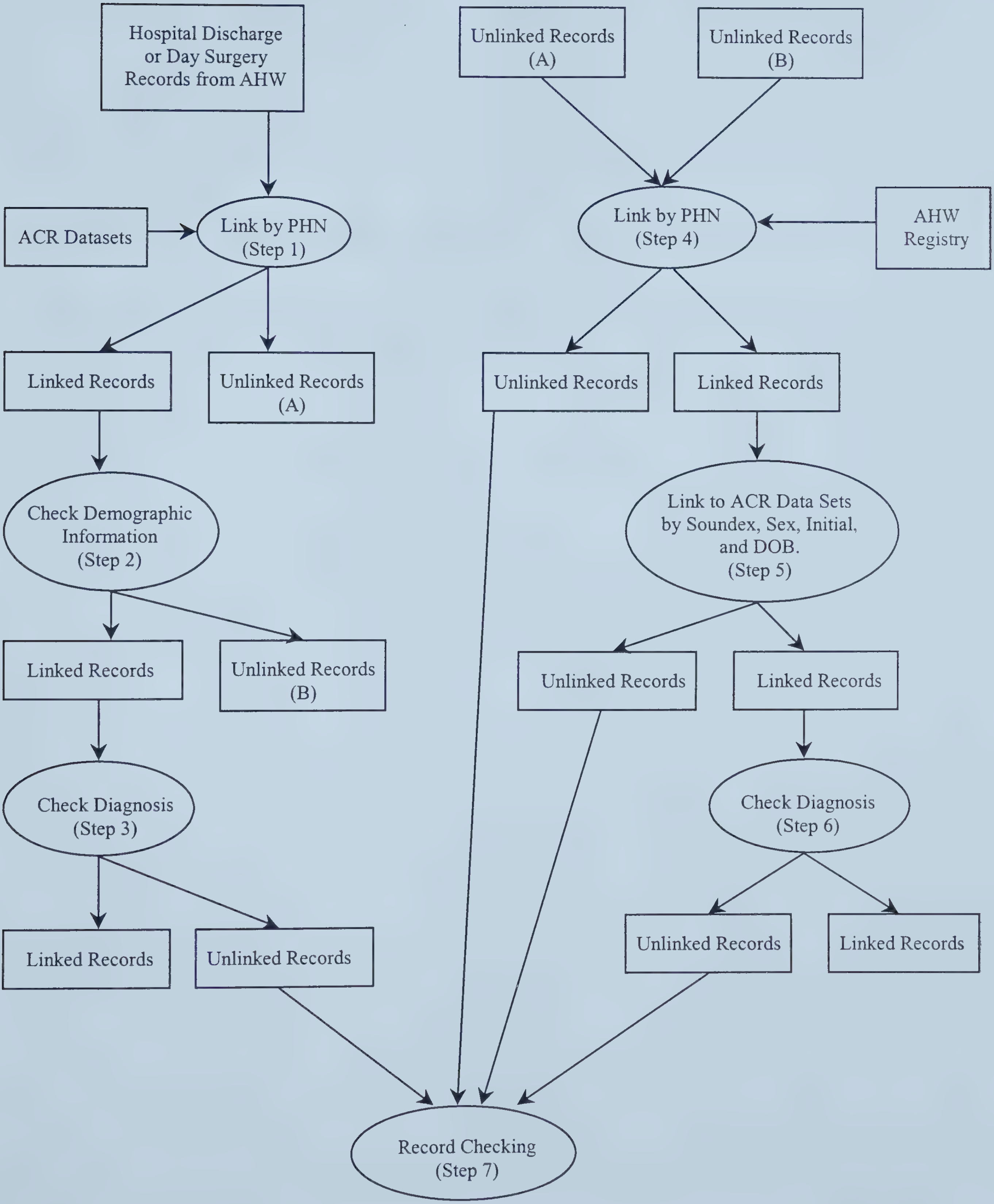


Table 4.1 Codes in Two Files Considered Equivalent[a]

ICD-9CM (AHW files)		ICD-O2 (ACR files)	
Codes	**Description**	**Codes**	**Description**
159	Malignant neoplasm of other and ill-defined sites within the digestive organs and peritoneum	C20	Rectum
173	Other malignant neoplasm of the skin	C00	Lip
179	Malignant neoplasm of the uterus, part unspecified	C53	Cervix uteri
180	Malignant neoplasm of the cervix uteri	C57	Other and unspecified female genital organs
204	Lymphoid leukemia	C77	Lymph nodes

[a] These are considered as extra to the standard conversion table.

B. The Incidence:Mortality (I:M) Ratio

1. **ACR Completeness Estimates**

The I:M ratio for a specific cancer is a comparison of the incidence number or rate with the mortality number or rate attributed to a given cancer in the same time-period. The I:M ratio can also be derived from the ratio of age-standardized rates. For the purposes of comparison, the illustration was made among crude rates, rates age-standardized to the 1991 Canadian population, and rates age-standardized to the 1970 U.S. population (Table 4.2). ACR completeness estimates using the I:M ratio were further derived based on the method used by NAACCR [72]. SEER incidence and NMDB mortality in the U.S. were used as a reference registry to calculate the expected incidence rate for the ACR. Completeness estimates were obtained through the comparison of the expected incidence rate with the actual ACR incidence rate on a site-specific basis. Results were compared with those published in the *CINA* by NAACCR.

Table 4.2 Comparisons of Crude and Standardized I:M Ratios in the ACR, 1994-96

		1994			1995			1996			1994-96		
		Crude	CDN Std.[a]	US Std.[b]	Crude	CDN Std.	US Std.	Crude	CDN Std.	US Std.	Crude	CDN Std.	US Std.
Female Breast	F	3.22	3.23	3.27	3.46	3.48	3.58	3.39	3.45	3.51	3.35	3.39	3.45
Prostate	M	4.49	4.21	4.46	3.68	3.44	3.67	4.42	4.12	4.41	4.17	3.90	4.16
Pancreas	M	1.11	1.10	1.09	1.03	1.01	1.02	1.11	1.11	1.10	1.08	1.07	1.07
	F	1.19	1.21	1.21	1.02	1.01	1.00	1.06	1.08	1.10	1.08	1.09	1.10
Melanoma of the Skin	M	5.50	5.11	4.97	7.22	6.31	6.33	5.81	5.70	5.45	6.12	5.68	5.56
	F	18.44	18.92	19.62	7.76	7.64	7.09	8.90	8.62	8.33	10.12	10.06	9.62
Colon and rectum	M	2.36	2.34	2.36	2.47	2.40	2.44	2.19	2.12	2.16	2.34	2.28	2.32
	F	2.44	2.49	2.51	2.30	2.37	2.44	2.15	2.18	2.19	2.29	2.34	2.37
Lung	M	1.24	1.23	1.24	1.15	1.13	1.15	1.20	1.19	1.20	1.20	1.18	1.19
	F	1.26	1.26	1.26	1.24	1.26	1.29	1.24	1.25	1.26	1.25	1.26	1.27
Leukemia	M	1.64	1.56	1.64	1.71	1.63	1.67	1.60	1.51	1.53	1.65	1.57	1.61
	F	1.50	1.51	1.56	1.97	1.96	2.04	1.48	1.51	1.64	1.64	1.64	1.74

[a] Age standardized to 1991 Canada population.

[b] Age standardized to 1970 U.S. population.

2. **Impacts of Different Rules on Completeness Estimates**

In order to investigate the implications of using different coding rules for multiple primaries on completeness estimates using I:M ratio, IARC rules were applied to ACR data and the corresponding I:M ratio with completeness estimates were compared to the results when SEER rules were used. With the 1970 U.S. census as standard population, the age-standardized SEER incidence rate and the age-standardized U.S. mortality rate were used to calculate the expected incidence rate for the ACR when SEER rules and IARC rules were used respectively. Completeness of the ACR was further derived by dividing the actual incidence rate in the ACR by the expected incidence rate above on a site-specific basis. Comparisons were made to explore the discrepancies between completeness estimates as a result of the different rules adopted.

C. Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%)

The percentage of diagnosed cases histologically confirmed was derived from the data field of "HIST_CNF" for the ACR on a site-specific basis. Similarly, the data field of "DIAG_MET" was used to derive the percentage of cases with a death certificate as the only source of cancer diagnosis for the ACR. The HV% and the DCO% were compared to the reference groups published by NAACCR. The possible reasons and factors contributing to the disparities in the comparison were investigated. Also, the comparison was made with previous years to explore the variations in registration completeness.

D. Case Re-finding in the University of Alberta Hospitals

UAH Outpatient records for the years 1994-96 were matched to ACR records to identify cases that failed to register in the ACR. In general, the procedures used to do the record linkage between Outpatient data from the UAH and the ACR data for the years 1994-96 were similar to those used in doing the record linkage between AHW and the ACR data. Unlike the AHW-ACR record linkage, the UAH case re-finding involved fewer steps because 1) Outpatient records from the UAH included all identifying information on cancer patients, and 2) multiple records existed for cases with multiple cancer diagnoses in the UAH files. With the advantage of the unique identifying variable "PHN" in both files, a deterministic record linkage between UAH and ACR data was conducted by performing the following steps:

STEP 1

Link data sets between ACR records and UAH records by the unique variable "PHN" coexisting in both files.

STEP 2

Check demographic information of date of birth and sex for those linked records in Step 1 to make sure those records are identical.

STEP 3

For those identical records in Step1, check their diagnostic information to see whether they agree with that recorded in the ACR.

STEP 4

For those records that were unable to be linked by "PHN" in Step 1 and those unlinked records in Step 2, link them back to the ACR data set by "SOUNDEX," "INITIAL," "SEX," and "DOB." Linked records were examined visually based on last name, first name, sex, and date of birth to make sure they were identical cases.

STEP 5

For those linked records in Step 4, check their diagnostic information to see whether they agreed with that recorded on the ACR

STEP 6

For those unlinked records in Step 4, as well as those records with diagnostic information disagreeing with that recorded in the ACR in Step 3 and in Step 5, case follow-back was undertaken by the ACR staff. After medical charts were examined, cases with questions concerning the diagnosis and those unregistered cases were followed back by writing letters to the University Hospital to confirm the diagnoses.

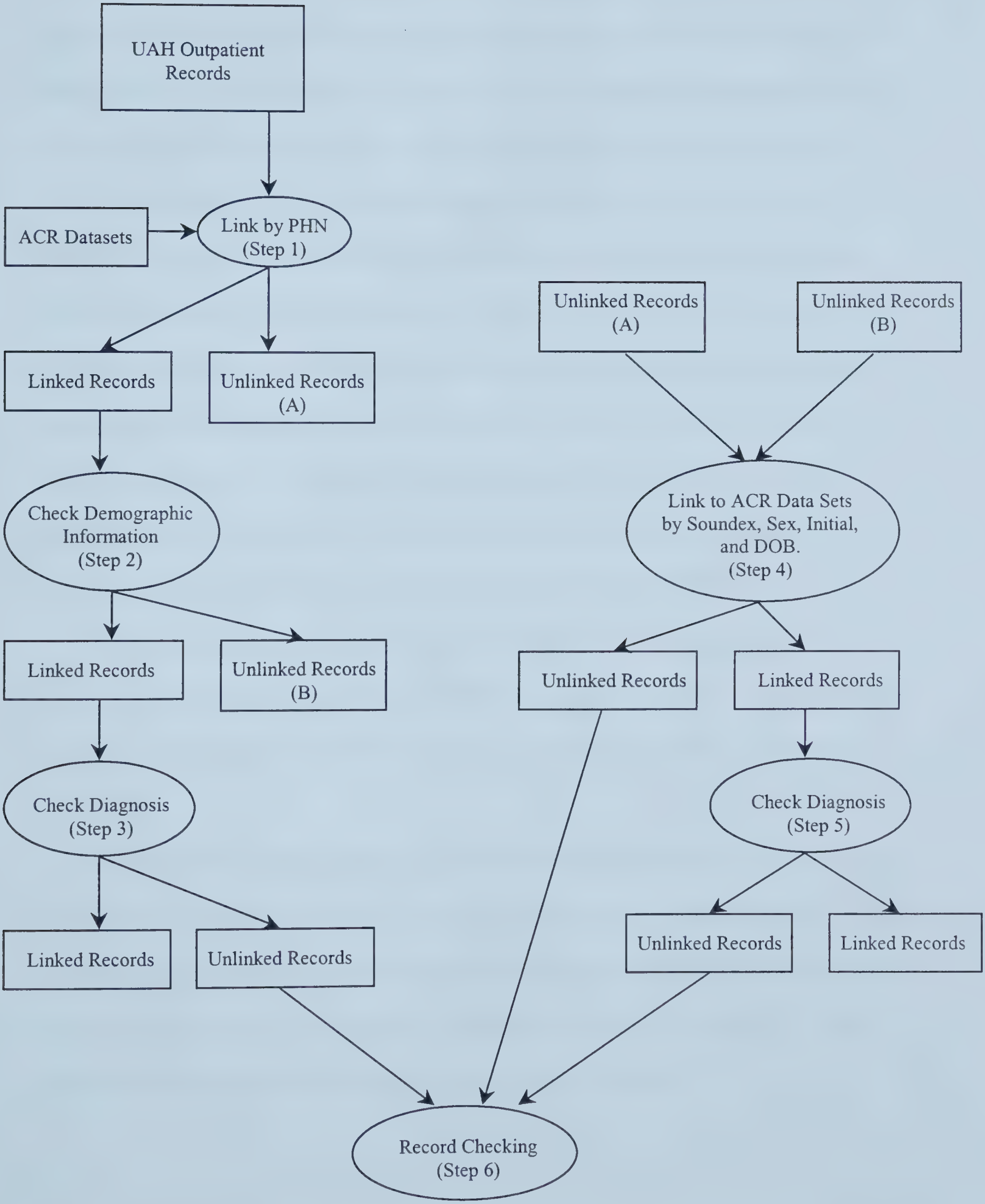
The same procedures were carried out to deal with differences in coding practice in UAH data (ICD-9CM) and in ACR data (ICD-O2) to check the agreement on diagnosis, along with codes considered equivalent in both data sets as those in the AHW-ACR linkage.

When the follow-back on unlinked cases is completed, cases that really are cancer cases but who did not appear in the ACR data will be registered. Additional information about the tumour will be added to the ACR for cases whose diagnoses did not agree

between the ACR and UAH data sets. After the results from the follow-back on unlinked cases are incorporated, the incompleteness of the ACR can be estimated by dividing the number of unlinked records from Step 3, Step 4, and Step 5 by the total number of original records in the UAH files for a certain period of time. Thus, completeness of the ACR can be estimated by levels of case coverage, which is defined as 1 – (unlinked UAH cancers/number of UAH cancers), and obtained as 1 – incompleteness. Because the UAH was the only hospital used in this case re-finding exercise, which included only a small proportion of all cases registered in the ACR, the number of records in the UAH data as the denominator for ACR completeness estimates was used. The case coverage has been reported in the results.

- Please refer to Figure 4.2 for this record linkage.

Figure 4.2 UAH Case Re-finding



E. Stratification by Demographic and Geographic Factors

In order to investigate the variation in completeness existing between the differences in sex, age group, and geographic region, the results of the AHW-ACR record linkage, I:M ratio, HV% and DCO%, and UAH case re-finding were further stratified by the potential factors contributing to completeness for comparison. In the AHW-ACR linkage and the UAH case re-finding, the degrees of completeness were estimated by dividing the number of AHW or UAH linked records in each factor-specific stratum by the number of records in the same stratum in the ACR or UAH files. For I:M ratio, HV%, and DCO% which used ACR data, completeness estimates were stratified by sex, age group, and geographic region based on information available in the ACR. Chi-square tests were used to test the significant differences among all completeness estimates, except for the I:M ratio, where the Chi-square test is not appropriate. For comparison, all major cancer sites except the female breast and prostate were stratified by sex.

For age comparison, age was defined as the number of completed years at first diagnosis/discharge/treated date. The age variable was grouped into five categories based on the rule used commonly in the health field; 0 - 14, 15 - 44, 45 - 64, 65 - 74, and older than 75 years.

For the comparison of geographic region, the postal codes of the patients' residences were classified as either rural or urban. The rules for distinguishing the rural/urban classification were based on the first two characters of the postal code. All Alberta postal codes start with the letter "T." If the first two characters of the postal code are "T0," the residence was considered rural; otherwise, it was considered urban.

F. Comparison of Methods

The completeness estimates using absolute measures of the AHW-ACR record linkage, I:M ratios, and UAH case re-finding were compared to results from relative measures of indicators of HV% and DCO%. Comparisons were also made of results from methods with absolute measures. On a site-specific basis, the total completeness estimates and estimates stratified by sex and region were compared. Because the information on regional stratification was not available for the standard registry of the SEER incidence and the NMDB mortality in the U.S., I:M ratios could be compared only with other methods based on sex stratification. In addition, the inclusion of cancers of the female breast and prostate in the overall completeness estimates using I:M ratios was investigated; these two cancers account for almost 30% of all cancer cases diagnosed in Alberta annually. Differences and variations in the comparisons were explored to find potential causes and explanations associated with the patterns.

CHAPTER V. RESULTS

A. The AHW-ACR Record Linkage

Overall, the case coverage of the ACR was 93.0% (95% CI: 92.76%, 93.24%) (Table 5.1). Among all the major cancer sites investigated, female breast cancer had the highest case coverage 98.3% (97.90%, 98.70%), followed by the cancer sites of the colon-rectum 93.9% (93.10%, 94.70%), lung 92.6% (91.79%, 93.41%), prostate 92.1% (91.29%, 92.91%), pancreas 91.6% (89.71%, 93.49%), melanoma of the skin 91.0% (89.30%, 92.70%), and leukemia 83.0% (80.65%, 85.35%).

The follow-back exercise is currently ongoing for unlinked cases and cases with diagnoses disagreeing with those recorded in the ACR. The up-to-date results of the case follow-back from this linkage are shown in Table 5.2. Results indicate that 805 out of 2,915 cases (27.6%) are benign cases, 284 cases (9.7%) have the same diagnosis, 489 cases (16.8%) are new cancer diagnoses, 148 cases (5.1%) have new primaries, and 1,016 cases (34.8%) are still waiting for replies from hospitals (i.e. 34% of unlinked cases that have been followed back are, in fact, cancers). Thus, the degree of completeness will be higher than the estimated case coverage after the case follow-back is complete. The overall ACR completeness is projected to be 97.6% after incorporating the most recent results from the case follow-back.

Table 5.1 Case Coverage[a] Estimated from the AHW-ACR Record Linkage, 1994-96

	Case Coverage % (N[b])	95% CI	
All Sites	93.0 (38967)	92.76	93.24
Female Breast	98.3 (3881)	97.90	98.70
Prostate	92.1 (3945)	91.29	92.91
Pancreas	91.6 (756)	89.71	93.49
Melanoma of the Skin	91.0 (986)	89.30	92.70
Colon-rectum	93.9 (3256)	93.10	94.70
Lung	92.6 (3749)	91.79	93.41
Leukemia	83.0 (817)	80.65	85.35

[a] Case coverage = 1 – [unlinked AHW cancers/(number of ACR cancers + unlinked AHW cancers)].

[b] Number of ACR cancers.

Table 5.2 Results of the Case Follow-back for the AHW-ACR Record Linkage

Status of unlinked cases or cases with disagreement on diagnosis	**Number of cases (%)**
Benign/ No Review	805 (27.6)
Same Diagnosis[a]	284 (9.7)
New Cancer Diagnosis	489 (16.8)
New Primary[b]	148 (5.1)
Diagnosis Questionable[c]	101 (3.5)
No Information[d]	72 (2.5)
No Reply from Hospitals	1016 (34.8)
Total	2915 (100.0)

[a] Considered same diagnosis after confirmation

[b] Any newly diagnosed primary cancer in a person previously diagnosed with cancer (i.e., not a metastatic cancer)

[c] Diagnosis can not be confirmed

[d] No other information can be found to confirm the diagnosis.

The case coverage for all sites was higher in females (93.5%) than in males (93.0%) (P=0.04) (Table 5.3). Although females appeared to have higher case coverage than males for cancer sites of the pancreas (92.6% vs 90.6%, P=0.30) and melanoma of the skin (91.3% vs 90.6%, P=0.66), these differences were not statistically significant. However, females had significantly lower case coverage than males for cancer sites of colon-rectum (92.9% vs 94.7%, P=0.03) and leukemia (79.4% vs 85.9%, P<0.01). Marginally significantly lower case coverage was found in females than in males for lung cancer (92.3% vs 93.7%, P=0.08).

When results of the record linkage were stratified by age group, the patterns were not consistent across cancer sites (Table 5.3). The degree of case coverage was at its highest in the age groups 15-44 years, 45-64 years, and 65-74 years, but dropped in the age group of older than 75 years for most sites investigated. The case coverage estimates in the first age group were unstable owing to small numbers. The Chi-square tests of age groups were statistically significant for all sites combined (P<0.01) and for cancers of the prostate (P<0.01), colon-rectum (P=0.04), lung (P<0.01), and leukemia (P<0.01).

Table 5.3 Results of the AHW-ACR Record Linkage Stratified by Sex, Age Group, and Region, 1994-96

	AB	Sex			Age Group[a]						Region		
	% (n^b)	**Male** % (n)	**Female** % (n)	**p**	**1** % (n)	**2** % (n)	**3** % (n)	**4** % (n)	**5** % (n)	**p**	**Rural** % (n)	**Urban** % (n)	**P**
All Sites	93.0 (38967)	93.0 (20708)	93.5 (18259)	0.04	71.9 (246)	91.4 (4328)	94.3 (12413)	94.2 (11180)	92.4 (10800)	<0.01	90.3 (9700)	94.1 (29267)	<0.01
Female Breast	98.3 (3881)		98.3 (3881)			98.1 (616)	98.4 (1692)	98.4 (858)	98.3 (715)	0.97	96.5 (834)	98.8 (3047)	<0.01
Prostate	92.1 (3945)	92.1 (3945)				42.9 (3)	94.3 (949)	94.3 (1733)	88.4 (1260)	<0.01	89.4 (1038)	93.1 (2907)	<0.01
Pancreas	91.6 (756)	90.6 (367)	92.6 (389)	0.30		84.8 (28)	90.9 (189)	91.3 (232)	93.0 (307)	0.39	87.8 (194)	93.0 (562)	0.02
Melanoma of the Skin	91.0 (986)	90.6 (470)	91.3 (516)	0.66		89.5 (366)	93.0 (343)	92.8 (155)	93.8 (121)	0.22	86.0 (190)	92.2 (796)	<0.01
Colon-rectum	93.9 (3256)	94.7 (1840)	92.9 (1416)	0.03		93.0 (133)	92.5 (931)	95.5 (1019)	93.8 (1173)	0.04	90.2 (847)	95.3 (2409)	<0.01
Lung	92.6 (3749)	93.7 (2288)	92.3 (1461)	0.08	50.0 (2)	82.9 (107)	94.7 (1205)	92.8 (1382)	91.6 (1053)	<0.01	90.6 (948)	93.3 (2801)	<0.01
Leukemia	83.0 (817)	85.9 (477)	79.4 (340)	<0.01	100.0 (69)	82.8 (111)	86.1 (216)	88.7 (197)	73.2 (224)	<0.01	77.6 (225)	85.3 (592)	<0.01

[a] Age Group - 1($0 \le$ age ≤ 14), 2($15 \le$ age ≤ 44), 3($45 \le$ age ≤ 64), 4($65 \le$ age ≤ 74), 5(age ≥ 75).

[b] Number of cancers in the ACR.

Among all major cancer sites, the pattern of urban regions having higher case coverage than rural regions was consistent (Table 5.3). Overall, the case coverage of the ACR was 94.1% and 90.3% ($P<0.01$) for urban and rural regions respectively. The equivalent figures for the urban and rural regions were 98.8% and 96.5% ($P<0.01$) for female breast, 93.1% and 89.4% ($P<0.01$) for prostate, 93.0% and 87.8% ($P=0.02$) for pancreas, 92.2% and 86.0% ($P<0.01$) for melanoma of the skin, 95.3% and 90.2% ($P<0.01$) for colon-rectum, 93.3% and 90.6% ($P<0.01$) for lung, and 85.3% and 77.6% ($P<0.01$) for leukemia.

B. The Incidence:Mortality (I:M) Ratio

1. **ACR Completeness Estimates**

The I:M ratios were calculated for major cancer sites for the years of 1994, 1995, 1996 individually, and for all three years combined, and were further stratified by sex, age group, and region of residence (Table 5.4) (Table 5.5). The comparisons of I:M ratios for different years on a site-specific basis did not show systematic changes in patterns. For the 1994-96 three-year period, I:M ratios for cancers of melanoma of the skin, prostate, and female breast were much higher than 1.0. Conversely, cancers of pancreas and lung had I:M ratios close to 1.0.

When the comparisons were made by sex, the patterns observed were not consistent across cancer sites. Melanoma of the skin and lung cancer were two sites in which females had higher I:M ratios than males across the three-year period. In particular, a much higher I:M ratio was observed for melanoma of the skin in females (9.92) than in males (6.03) (Table 5.5). When the comparisons were made among age

groups, people in older age groups tended to have lower I:M ratios than people in younger age groups. The trends were consistent across all major cancer sites. In the comparison of region of residence, people living in urban regions had lower I:M ratios than people living in rural regions for all major cancer sites, except for prostate and melanoma of the skin.

Table 5.4 Crude I:M Ratios for ACR Completeness by Year for 1994-96

Cancer Sites	1994-96	1994	1995	1996
Female Breast	3.33	3.19	3.46	3.40
Prostate	4.16	4.49	3.68	4.40
Pancreas	1.07	1.13	1.01	1.07
Melanoma of the Skin	7.62	8.62	7.39	7.08
Colon-rectum	2.28	2.36	2.39	2.13
Lung	1.21	1.25	1.18	1.21
Leukemia	1.64	1.58	1.78	1.56

Table 5.5 Crude I:M Ratios for ACR Completeness by Sex, Age Group, and Region, 1994-96

Cancer Sites	AB	Sex		Age Group[a]					Region	
		Male	Female	1	2	3	4	5	Rural	Urban
Female Breast	3.33		3.33		4.50	4.13	2.95	2.13	3.89	3.20
Prostate	4.16	4.16			3.01	10.50	6.37	2.14	3.78	4.32
Pancreas	1.07	1.07	1.07		1.67	1.05	1.16	0.99	1.28	1.01
Melanoma of the Skin	7.62	6.03	9.92		23.82	6.58	5.12	3.91	6.95	7.79
Colon-rectum	2.28	2.30	2.25		4.66	2.50	2.61	1.84	2.79	2.15
Lung	1.21	1.19	1.24		1.48	1.34	1.23	1.06	1.49	1.14
Leukemia	1.64	1.65	1.62	4.66	1.90	1.81	1.80	1.10	1.97	1.54

[a] AgeGroup-1($0 \le age \le 14$),2($15 \le age \le 44$),3($45 \le age \le 64$),4($65 \le age \le 74$),5($age \ge 75$).

The overall ACR completeness for the years 1994-96 estimated by I:M ratio using the NAACCR methods were 88.8% for males (including prostate cancer), 84.6% for females (including breast cancer), and 86.9% for both sexes combined (Table 5.7). However, the ACR completeness estimates were 96.4% for males (excluding prostate cancer), 98.3% for females (excluding breast cancer), and 97.2% for both sexes combined. Among the major sites investigated, cancers of the pancreas, lung, and leukemia had completeness estimates higher than 100% in both males and females. Much lower completeness estimates using I:M ratios were observed for cancers of the female breast (65.8%) and prostate (75.3%) than for the other major sites investigated. When the comparisons were made by sex, females had higher levels of completeness than males in the sites of pancreas (108.9% vs 105.6%), melanoma of the skin (92.7% vs 72.0%), colon-rectum (90.2% vs 87.7%), and leukemia (107.0% vs 100.8%), but not for lung (104.5% vs 118.4%).

2. Impacts of Different Rules on Completeness Estimates

When the ACR completeness estimates were made using I:M ratios with different coding rules, the number of incident cases using SEER rules was greater than or equal to the number using IARC rules in all sites (Table 5.6). The largest differences among males were observed for colon-rectum (44 cases (2.3%)), urinary bladder (41 cases (4.3%)), and lung & bronchus (14 cases (0.6%). For females, the largest differences in incidence were observed for breast (187 cases (4.8%)), colon-rectum (48 cases (3.2%)), and melanoma of the skin (27 cases (5.2%)). These differences resulted in a maximum absolute difference in the I:M ratio of 0.52 for female melanoma of the skin.

The difference in the completeness estimates using the two methods for coding was relatively small for most sites (Table 5.7). The largest differences were for female melanoma of the skin (5.0%), male urinary bladder (4.6%), male Hodgkin's disease (3.6%), and female breast (3.2%). It should be noted that the differences in completeness estimates between the two coding rules were small whether prostate and female breast cancers were included. However, the differences in the degree of overall completeness estimates were not negligible when prostate and female breast cancers were included. The difference in completeness estimates for all sites between the two coding methods was 1.3% with prostate and female breast included, and 1.2% with them excluded.

Table 5.6 Differences in Incidence and Crude I:M Ratios as a Result of the Adoption of Different Coding Rules in the ACR, 1994-96

Males **Cancer Site**	**SEER Incidence**	**IARC Incidence**	**Incidence Difference (%)**	**Mortality**	**Crude SEER I:M Ratio**	**Crude IARC I:M Ratio**	**I:M Ratio Difference**
Oral cavity & pharynx	510	502	8 (1.57)	152	3.36	3.30	0.06
Esophagus	146	146	0	110	1.33	1.33	0.00
Stomach	376	375	1 (0.27)	302	1.25	1.24	0.01
Colon-rectum	1901	1857	44 (2.31)	825	2.30	2.25	0.05
Liver	166	166	0	129	1.29	1.29	0.00
Pancreas	367	367	0	344	1.07	1.07	0.00
Lung & bronchus	2291	2277	14 (0.61)	1931	1.19	1.18	0.01
Melanoma of the skin	470	458	12 (2.55)	78	6.03	5.87	0.16
Prostate	3932	3931	1 (0.03)	945	4.16	4.16	0.00
Urinary bladder	949	908	41 (4.32)	179	5.30	5.07	0.23
Kidney & renal pelvis	537	534	3 (0.56)	190	2.83	2.81	0.02
Brain & CNS	281	280	1 (0.36)	225	1.25	1.24	0.01
Hodgkin's disease	111	108	3 (2.70)	17	6.53	6.35	0.18
NHL	572	570	2 (0.35)	278	2.06	2.05	0.01
Myelomas	166	166	0	128	1.30	1.30	0.00
Leukemias	473	473	0	287	1.65	1.65	0.00

Females **Cancer Site**	**SEER Incidence**	**IARC Incidence**	**Incidence Difference (%)**	**Mortality**	**Crude SEER I:M Ratio**	**Crude IARC I:M Ratio**	**I:M Ratio Difference**
Oral cavity & pharynx	201	201	0	73	2.75	2.75	0.00
Esophagus	71	71	0	56	1.27	1.27	0.00
Stomach	224	224	0	185	1.21	1.21	0.00
Colon-rectum	1480	1432	48 (3.24)	658	2.25	2.18	0.07
Liver	77	77	0	71	1.08	1.08	0.00
Pancreas	389	389	0	362	1.07	1.07	0.00
Lung & bronchus	1461	1454	7 (0.48)	1182	1.24	1.23	0.01
Melanoma of the skin	516	489	27 (5.23)	52	9.92	9.40	0.52
Breast (excl. *in situ*)	3872	3685	187 (4.83)	1164	3.33	3.17	0.16
Cervix – invasive	392	392	0	126	3.11	3.11	0.00
Corpus & uterus, NOS	797	797	0	129	6.18	6.18	0.00
Ovary	570	570	0	292	1.95	1.95	0.00
Urinary bladder	318	311	7 (2.20)	66	4.82	4.71	0.11
Kidney & renal pelvis	329	327	2 (0.61)	103	3.19	3.17	0.02
Brain & CNS	202	202	0	146	1.38	1.38	0.00
Hodgkin's disease	109	108	1 (0.92)	9	12.11	12.00	0.11
NHL	463	460	3 (0.65)	224	2.07	2.05	0.02
Myelomas	140	140	0	107	1.31	1.31	0.00
Leukemias	338	337	1 (0.30)	209	1.62	1.61	0.01

Table 5.7 Comparisons of ACR Completeness Estimates Using I:M Ratios between the Adoption of SEER and IARC Rules, 1994-96

Males **Cancer Site**	**ACR SEER Inc Rate**[a]	**ACR IARC Inc Rate**[a]	**ACR Mrt Rate**[a]	**SEER I:M Ratio**[b]	**Exp Inc Rate**[c]	**SEER Complete**[d] **(%)**	**IARC Complete**[e] **(%)**	**Difference In Complete (%)**
Oral cavity & pharynx	13.63	13.41	4.06	3.8	15.43	88.3	86.9	1.4
Esophagus	4.06	4.06	3.05	1.0	3.05	133.1	133.1	0.0
Stomach	10.44	10.41	8.44	1.9	16.04	65.1	64.9	0.2
Colon-rectum	52.80	51.56	23.15	2.6	60.19	87.7	85.7	2.0
Liver	4.52	4.52	3.55	1.4	4.97	90.9	90.9	0.0
Pancreas	10.18	10.18	9.64	1.0	9.64	105.6	105.6	0.0
Lung & bronchus	64.29	63.89	54.31	1.0	54.31	118.4	117.6	0.8
Melanoma of the skin	11.39	11.06	2.08	7.6	15.81	72.0	70.0	2.0
Prostate	111.67	111.64	26.95	5.5	148.23	75.3	75.3	0.0
Urinary bladder	26.28	25.13	5.05	5.1	25.76	102.2	97.6	4.6
Kidney & renal pelvis	14.31	14.22	5.26	2.5	13.15	108.8	108.1	0.7
Brain & CNS	7.13	7.10	5.73	1.3	7.45	95.7	95.3	0.4
Hodgkin's disease	2.72	2.65	0.42	4.6	1.93	140.9	137.3	3.6
NHL	14.82	14.77	7.63	2.4	18.31	80.9	80.7	0.2
Myelomas	4.54	4.54	3.57	1.3	4.64	97.8	97.8	0.0
Leukemias	12.74	12.74	7.90	1.6	12.64	100.8	100.8	0.0
Total Males	**365.52**	**361.88**			**411.55**	**88.8**	**87.9**	**0.9**
Total Males (excl. prostate)	**253.85**	**250.24**			**263.32**	**96.4**	**95.0**	**1.4**
Females **Cancer Site**	**ACR SEER Inc Rate**[a]	**ACR IARC Inc Rate**[a]	**ACR Mrt Rate**[a]	**SEER I:M Ratio**[b]	**Exp Inc Rate**[c]	**SEER Complete**[d] **(%)**	**IARC Complete**[e] **(%)**	**Difference In Complete (%)**
Oral cavity & pharynx	4.70	4.70	1.73	4.6	7.96	59.0	59.0	0.0
Esophagus	1.54	1.54	1.20	1.2	1.44	106.9	106.9	0.0
Stomach	4.81	4.81	3.88	1.7	6.60	72.9	72.9	0.0
Colon-rectum	33.39	32.37	14.23	2.6	37.00	90.2	87.5	2.7
Liver	1.79	1.79	1.61	1.4	2.25	79.6	79.6	0.0
Pancreas	8.59	8.59	7.89	1.0	7.89	108.9	108.9	0.0
Lung & bronchus	35.98	35.81	28.70	1.2	34.44	104.5	104.0	0.5
Melanoma of the skin	11.16	10.56	1.18	10.2	12.04	92.7	87.7	5.0
Breast (excl. *in situ*)	91.58	87.14	26.78	5.2	139.26	65.8	62.6	3.2
Cervix – invasive	7.96	7.96	2.73	3.0	8.19	97.2	97.2	0.0
Corpus & uterus, NOS	19.61	19.61	2.88	6.4	18.43	106.4	106.4	0.0
Ovary	12.99	12.99	6.89	1.9	13.09	99.2	99.2	0.0
Urinary bladder	7.38	7.25	1.39	4.3	5.98	123.4	121.2	2.2
Kidney & renal pelvis	7.80	7.75	2.40	2.7	6.48	120.4	119.6	0.8
Brain & CNS	4.80	4.80	3.54	1.4	4.96	96.8	96.8	0.0
Hodgkin's disease	2.54	2.51	0.20	7.4	1.48	171.6	169.6	2.0
NHL	10.74	10.66	5.02	2.2	11.04	97.3	96.6	0.7
Myelomas	3.27	3.27	2.48	1.4	3.47	94.2	94.2	0.0
Leukemias	7.93	7.91	4.63	1.6	7.41	107.0	106.7	0.3
Total Females	**278.56**	**272.02**			**329.41**	**84.6**	**82.6**	**2.0**
Total Females (excl. breast)	**186.98**	**184.88**			**190.15**	**98.3**	**97.2**	**1.1**
Overall Totals	**644.08**	**633.90**			**740.96**	**86.9**	**85.6**	**1.3**
Overall Totals (excl. prostate and breast)	**440.83**	**435.12**			**453.47**	**97.2**	**96.0**	**1.2**

[a] Age standardized to 1970 U.S. population.

[b] (SEER 1994-96 incidence rate) / (U.S. 1994-96 mortality rate).

[c] (ACR mortality rate) X (SEER I:M ratio).

[d] Completeness of ACR estimated by SEER rules.

[e] Completeness of ACR estimated by IARC rules.

C. Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%) as Completeness Indicators

Completeness indicators of HV% and DCO% were derived from ACR data for the years 1994, 1995, and 1996, as well as the three years combined (Table 5.8). Results were stratified by sex, age group, and region of residence (Table 5.9). The comparison of HV% and DCO% among three-year periods did not show patterns of systematic change on a site-specific basis. The indicators of HV% varied across sites. Melanoma of the skin had the highest HV% (100%) among sites investigated, followed by leukemia (98.3%), female breast (98.2%), colon-rectum (95.6%), prostate (94.9%), and lung (87.7%), while pancreas had the lowest HV% (56.2%).

In general, females appeared to have lower HV% than males (Table 5.9). However, a significant difference was only found for pancreas (53.0% vs 59.7%, $P=0.04$). People in the younger age groups tended to have a higher HV% than people in the older age groups. The Chi-square tests of age groups were statistically significant for all major cancer sites investigated except for melanoma of the skin.

People living in urban regions tended to have higher HV% as opposed to people living in rural regions. Chi-square tests of comparisons between urban and rural regions showed statistically significant differences for cancer sites of the female breast ($P<0.01$), prostate ($P<0.01$), colon-rectum ($P=0.02$), and leukemia ($P=0.04$).

Table 5.8 HV% and DCO% as ACR Completeness Indicators by Year, 1994-96

Cancer Sites		1994-96	1994	1995	1996
Female Breast					
	HV%	98.2	98.5	98.0	98.2
	DCO%	0.1	0.0	0.2	0.1
Prostate					
	HV%	94.9	94.4	94.6	95.6
	DCO%	0.1	0.1	0.1	0.1
Pancreas					
	HV%	56.2	50.2	53.9	63.2
	DCO%	0.7	0.0	0.4	1.4
Melanoma of the Skin					
	HV%	100.0	100.0	100.0	100.0
	DCO%	0.0	0.0	0.0	0.0
Colon-rectum					
	HV%	95.6	95.2	96.5	95.2
	DCO%	0.1	0.1	0.0	0.2
Lung					
	HV%	87.7	88.3	88.7	86.2
	DCO%	0.1	0.2	0.0	0.2
Leukemia					
	HV%	98.3	98.7	97.9	98.9
	DCO%	0.4	0.4	0.4	0.4

Table 5.9 HV% and DCO% as ACR Completeness Indicators by Sex, Age Group, and Region, 1994-96

		AB	Sex			Age Group[a]						Region		
			Male	Female	p	1	2	3	4	5	p	Rural	Urban	P
Female Breast	HV%	98.2		98.2			99.5	99.6	98.9	93.4	<0.01	97.2	98.5	<0.01
	DCO%	0.1		0.1			0.0	0.1	0.1	0.1	0.78	0.2	0.0	0.06
Prostate	HV%	94.9	94.9				100.0	99.2	98.0	86.9	<0.01	93.2	95.5	<0.01
	DCO%	0.1	0.1				0.0	0.0	0.0	0.3	<0.01	0.2	0.1	0.28
Pancreas	HV%	56.2	59.7	53.0	0.04		96.3	76.6	61.7	35.9	<0.01	58.9	55.3	0.42
	DCO%	0.7	0.6	0.8	0.70		0.0	0.0	0.9	1.0	0.59	0.5	0.7	0.77
Melanoma of the Skin	HV%	100.0	100.0	100.0		100.0	100.0	100	100.0	100.0		100.0	100.0	
	DCO%	0.0	0.0	0.0		0.0	0.0	0.0	0.0	0.0		0.0	0.0	
Colon-Rectum	HV%	95.6	96.0	95.2	0.13		100.0	98.5	97.0	91.6	<0.01	94.3	96.1	0.02
	DCO%	0.1	0.1	0.2	0.42		0.0	0.0	0.1	0.2	0.67	0.1	0.1	0.77
Lung	HV%	87.7	87.4	88.0	0.37	100.0	98.1	96.2	90.5	72.8	<0.01	86.2	88.2	0.12
	DCO%	0.1	0.2	0.1	0.26	0.0	0.0	0.0	0.1	0.5	0.03	0.2	0.1	0.65
Leukemia	HV%	98.3	98.7	97.8	0.13	100.0	100.0	100.0	98.5	95.6	<0.01	98.1	98.4	0.04
	DCO%	0.4	0.2	0.6	0.40	0.0	0.0	0.0	0.0	1.5	<0.01	0.9	0.2	0.01

[a] AgeGroup-1($0 \le age \le 14$),2($15 \le age \le 44$),3($45 \le age \le 64$),4($65 \le age \le 74$),5($age \ge 75$).

When using DCO% as completeness indicators, the patterns observed for sex, age group, and region comparisons were consistent with those observed when using HV% as indicators (Table 5.9). The patterns of completeness indicators of DCO% being higher in older people were consistent across cancer sites except for melanoma of the skin. Chi-square tests of age groups were statistically significant for cancers of the prostate ($P<0.01$), lung ($P=0.03$) and leukemia ($P<0.01$). People living in urban regions tended to have a lower DCO% than people living in rural regions among sites investigated except for the pancreas. However, a significant difference was found only for leukemia in this comparison ($P=0.01$). Accordingly, the fact that people in older age groups and people living in rural regions had a higher DCO% and a lower HV% indicated the lower levels of case coverage.

D. The UAH Case Re-finding

Overall, the result of this record linkage was that 16 out of the 16,383 records (0.1%) had a disagreement on diagnosis, and 732 out of the 16,383 records (4.5%) did not appear in the ACR database, for a total of unlinked records of 748 out of 16,383, with a case coverage estimate of 95.4% (95.08%, 95.72%) (Table 5.10). Among major cancer sites investigated, leukemia had the highest percentage of records matched 99.0% (98.56%, 99.44%), followed by female breast 97.7% (96.55%, 98.85%), prostate 96.5% (95.32%, 97.68%), colon-rectum 95.7% (94.02%, 97.38), lung 94.3% (92.79%, 95.81%), melanoma of the skin 90.5% (89.06%, 91.94%), and pancreas 73.3% (62.11%, 84.49%). Among these cancer sites, it appears that leukemia had the highest percentage of case

coverage in the case re-finding, whereas pancreas had the lowest. Currently, the process of case follow-back is still ongoing; preliminary results indicate that the final estimates of completeness will be higher than the case coverage estimates.

Table 5.10 Case Coverage[a] Estimated from the UAH Case Re-finding, 1994-96

	Case Coverage % (N[b])	95% CI	
All Sites	95.4 (16383)	95.08	95.72
Female Breast	97.7 (645)	96.55	98.85
Prostate	96.5 (935)	95.32	97.68
Pancreas	73.3 (60)	62.11	84.49
Melanoma of the Skin	90.5 (1584)	89.06	91.94
Colon-rectum	95.7 (560)	94.02	97.38
Lung	94.3 (903)	92.79	95.81
Leukemia	99.0 (1954)	98.56	99.44

[a] Case coverage = 1 – (unlinked UAH cancers/number of UAH cancers).

[b] Number of UAH cancers.

The results of this case re-finding were further stratified by sex, age group, and region to investigate the impacts that these factors have on case coverage estimates (Table 5.11). For the comparison of sex, the patterns observed were not consistent across sites. Females had a significantly lower case coverage than males with all sites combined (95.0% vs 95.8%, P=0.02), and for colon-rectum cancer (93.8% vs 97.6%, P=0.03). Females had a significantly higher case coverage than males for leukemia (99.4% vs 98.3%, P=0.01). Although a higher case coverage in females than males was also found in cancers of the pancreas, the difference was not statistically significant (86.4% vs 65.8%, P=0.08).

When the comparisons were made by age group, more complete coverage was achieved both in the oldest and the youngest age groups for all sites combined (P<0.01) (Table 5.11). However, the patterns were not consistent across sites. For cancers of colon-rectum and lung, people in the oldest age groups appeared to have lower case coverage than people in the younger age groups. Except for female breast cancers (P=0.07), the Chi-square tests were statistically significant for the other major sites investigated.

When the results were stratified by region of residence, urban regions had a case coverage similar to rural regions with all sites combined (95.5% vs 95.2%, P=0.56). However, the results varied by site. In general, the comparisons between urban and rural regions were not statistically significantly different. Urban regions tended to have higher case coverage than rural regions for cancers of the female breast (98.0% vs 96.5%, P=0.29), colon-rectum (95.8% vs 95.5%, P=0.88), and lung (95.4% vs 92.3%, P=0.06), but urban regions tended to have a lower degree of completeness for cancers of the

prostate (96.2% vs 97.6%, P=0.32), pancreas (70.4% vs 100.0%, P=0.12), and melanoma of the skin (90.0% vs 93.4%, P=0.10). However, leukemia was the only site for which urban regions had significantly lower case coverage than rural regions (98.8% vs 100.0%, P=0.02).

Table 5.11 Results of UAH Case Re-finding Stratified by Sex, Age Group, and Region, 1994-96

	AB	Sex			Age Group[a]						Region		
	% (n[b])	Male % (n)	Female % (n)	p	1 % (n)	2 % (n)	3 % (n)	4 % (n)	5 % (n)	p	Rural % (n)	Urban % (n)	P
All Sites	95.4 (16383)	95.8 (8061)	95.0 (8318)	0.02	96.8 (1378)	95.9 (4846)	94.4 (5131)	95.0 (3231)	96.8 (1794)	<0.01	95.2 (3043)	95.5 (13340)	0.56
Female Breast	97.7 (645)		97.7 (645)			96.8 (93)	98.4 (308)	95.3 (150)	100 (97)	0.07	96.5 (143)	98.0 (505)	0.29
Prostate	96.5 (935)	96.5 (935)			100 (2)	25.0 (4)	94.9 (274)	97.3 (413)	97.9 (243)	<0.01	97.6 (208)	96.2 (728)	0.32
Pancreas	73.3 (60)	65.8 (38)	86.4 (22)	0.08		88.9 (9)	45.0 (20)	66.7 (9)	95.5 (22)	<0.01	100 (6)	70.4 (54)	0.12
Melanoma of the Skin	90.5 (1584)	91.2 (646)	90.0 (938)	0.43	7.7 (13)	87.1 (597)	92.4 (596)	93.9 (262)	99.1 (116)	<0.01	93.4 (229)	90.0 (1355)	0.10
Colon-rectum	95.7 (560)	97.6 (286)	93.8 (274)	0.03		89.5 (19)	98.2 (232)	94.9 (197)	92.9 (112)	0.05	95.5 (133)	95.8 (427)	0.88
Lung	94.3 (903)	95.4 (501)	93.0 (402)	0.12	100 (3)	85.7 (56)	93.9 (354)	96.1 (362)	94.2 (137)	0.04	92.3 (300)	95.4 (603)	0.06
Leukemia	99.0 (1954)	98.3 (703)	99.4 (1251)	0.01	96.4 (112)	99.3 (741)	99.1 (694)	99.5 (202)	98.5 (205)	0.05	100 (430)	98.8 (1524)	0.02

[a] Age Group - 1($0 \leq age \leq 14$), 2($15 \leq age \leq 44$), 3($45 \leq age \leq 64$), 4($65 \leq age \leq 74$), 5($age \geq 75$).

[b] Number of UAH cancers.

E. Comparison of Methods

In all the above methods used to estimate ACR completeness, the results from the AHW-ACR record linkage, I:M ratio, and UAH case re-finding gave absolute estimates, and the results from HV% and DCO% gave relative indicators. The results from UAH case re-finding as an independent case ascertainment method had a higher case coverage estimate (95.4%) than the results from the AHW-ACR record linkage (93.0%). The I:M ratio gave an estimate of 86.9% for completeness with the inclusion of female breast and prostate cancers which was lower than the two case coverage estimates (Table 5.12). However, the results from I:M ratios had the highest completeness estimate (97.2%) among methods giving absolute measures when female breast and prostate cancers were excluded. The ACR completeness estimate from the AHW-ACR record linkage and UAH case re-finding are expected to be higher after incorporating the results from the case follow-back.

When comparisons were made for completeness estimates between absolute and relative measures, the indicators of HV% and DCO% showed the relatively high completeness for melanoma of the skin (HV%: 100.0% and DCO%: 0.0%), whereas pancreas cancers had relatively low completeness (HV%: 56.2% and DCO%: 0.7%) among the major sites investigated (Table 5.12). Contrarily, results from the AHW-ACR record linkage showed the highest case coverage for female breast cancers (98.3%) but the lowest case coverage for leukemia (83.0%). Leukemia had the highest case coverage (99.0%) and pancreatic cancers had the lowest (73.3%) in the UAH case re-finding.

For all sites combined, females had a lower degree of case coverage than males in the results from the UAH case re-finding and I:M ratios, but not in the AHW-ACR record

linkage (Table 5.12). The pattern that females had higher case coverage than males in pancreatic cancers was consistent across methods with absolute measures, but not in relative measures of HV% and DCO%. Conversely, the pattern that females had lower case coverage than males in lung cancers was consistent across methods with absolute measures, but not in relative measures of HV% and DCO%. Among the three methods giving absolute measures of completeness estimates, results from the I:M ratio were much lower than those of others for cancers of the female breast (65.8%) and prostate (75.3%). The largest difference in completeness estimates in sex comparison was found for melanoma of the skin using the I:M ratio method (female 92.7% vs male 72.0%).

Table 5.12 Comparisons of ACR Completeness Estimates by Different Methods with Sex Stratification, 1994-96

	Absolute Measures									Relative Measures		
	AHW-ACR Record Linkage (%)			I:M Ratio[a]			UAH Case Re-Finding (%)			HV%, DCO%		
	AB	Male	Female	AB	Male	Female	AB	Male	Female	AB	Male	Female
All Sites	93.0	93.0	93.5	86.9	88.8	84.6	95.4	95.8	95.0			
Female Breast	98.3		98.3			65.8	97.7		97.7	98.2, 0.1		98.2, 0.1
Prostate	92.1	92.1			75.3		96.5	96.5		94.9, 0.1	94.9, 0.1	
Pancreas	91.6	90.6	92.6		105.6	108.9	73.3	65.8	86.4	56.2, 0.7	59.7, 0.6	53.0, 0.8
Melanoma of the Skin	91.0	90.6	91.3		72.0	92.7	90.5	91.2	90.0	100.0, 0.0	100.0, 0	100.0, 0.0
Colon-rectum	93.9	94.7	92.9		87.7	90.2	95.7	97.6	93.8	95.6, 0.1	96.0, 0.1	95.2, 0.2
Lung	92.6	93.7	92.3		118.4	104.5	94.3	95.4	93.0	87.7, 0.1	87.4, 0.2	88.0, 0.1
Leukemia	83.0	85.9	79.4		100.8	107.0	99.0	98.3	99.4	98.3, 0.4	98.7, 0.2	97.8, 0.6

[a] With the inclusion of female breast and prostate cancers.

Although the pattern that people living in urban regions had significantly higher case coverage than in rural regions was found in results from the AHW-ACR record linkage (94.1% vs 90.3%, P<0.01) with all sites combined, the results from the UAH case re-finding did not show a significant difference (95.5% vs 95.2%, P=0.56) (Table 5.13). Although urban regions had a significantly higher case coverage than rural regions for leukemia among the major sites investigated in the AHW-ACR record linkage (P<0.01), urban regions had significantly lower case coverage than rural regions for leukemia in the UAH case re-finding (P=0.02). However, significantly higher case coverage and marginally significantly higher case coverage in urban regions than in rural regions were found for lung cancer in the AHW-ACR record linkage (P<0.01) and the UAH case re-finding (P=0.06) respectively. The patterns on the regional comparison among methods were not consistent when the results were compared on a site-specific basis. Urban regions had significantly higher case coverage than rural regions in results from both AHW-ACR record linkage and relative measures for cancers of the female breast, prostate, colon-rectum, and leukemia. The largest difference in completeness estimates in the regional comparison was found for pancreatic cancers in the UAH case re-finding (urban 70.4% vs rural 100.0%).

Table 5.13 Comparisons of ACR Completeness Estimates by Different Methods with Regional Stratification, 1994-96

	Absolute Measures								Relative Measures			
	AHW-ACR Record Linkage (%)				UAH Case Re-finding (%)				HV%, DCO%			
	AB	Rural	Urban	P	AB	Rural	Urban	P	AB	Rural	Urban	P
All Sites	93.0	90.3	94.1	<0.01	95.4	95.2	95.5	0.56				
Female Breast	98.3	96.5	98.8	<0.01	97.7	96.5	98.0	0.29	98.2, 0.1	97.2, 0.2	98.5, 0.0	<0.01, 0.06
Prostate	92.1	89.4	93.1	<0.01	96.5	97.6	96.2	0.32	94.9, 0.1	93.2, 0.2	95.5, 0.1	<0.01, 0.28
Pancreas	91.6	87.8	93.0	0.02	73.3	100	70.4	0.12	56.2, 0.7	58.9, 0.5	55.3, 0.7	0.42, 0.77
Melanoma of the Skin	91.0	86.0	92.2	<0.01	90.5	93.4	90.0	0.10	100.0, 0.0	100.0, 0.0	100.0, 0.0	
Colon-rectum	93.9	90.2	95.3	<0.01	95.7	95.5	95.8	0.88	95.6, 0.1	94.3, 0.1	96.1, 0.1	0.02, 0.77
Lung	92.6	90.6	93.3	<0.01	94.3	92.3	95.4	0.06	87.7, 0.1	86.2, 0.2	88.2, 0.1	0.12, 0.65
Leukemia	83.0	77.6	85.3	<0.01	99.0	100	98.8	0.02	98.3, 0.4	98.1, 0.9	98.4, 0.2	0.04, 0.01

F. Summary

Overall, ACR case coverage for the years 1994-96, estimated by independent case ascertainment methods from the AHW-ACR record linkage, was 93.0%, expected to provide an estimate of completeness of around 97.6% after incorporating the results from the case follow-back. Cancers of the female breast and leukemia were estimated by this method to have the highest and the lowest case coverage respectively.

ACR completeness estimates varied depending on methods used. Variations also existed when results were stratified by factors of sex, age group, and region of residence. In general, females tended to have higher case coverage than males in the results from the AHW-ACR record linkage with all sites combined, but not for the UAH case re-finding and I:M ratio method. The pattern that older people tended to have lower case coverage than younger people was consistent across all methods used. The pattern that people living in urban regions tended to have higher case coverage than those in rural regions was observed consistently across methods except for the I:M ratio method.

In sum, results of the ACR completeness estimates varied based on the method used. Results from relative measures should be considered together with the results from absolute measures when interpreting ACR completeness estimates.

CHAPTER VI. DISCUSSION

In this thesis, four different methods of measuring the completeness of case ascertainment of the ACR have been presented. This chapter contains a discussion of the methods used and how their attributes may affect the estimates of ACR completeness.

A. ACR Completeness Estimates

1. **The AHW-ACR Record Linkage**

Although AHW files used in this record linkage were considered comprehensive, it should be noted that only inpatient records of hospital discharge and day surgery cases were included in the AHW files. Diagnoses for cancer patients might not be confirmed when they are discharged. In addition, some of the unlinked cases with a disagreement on diagnosis may have resulted from the slight discrepancies in diagnoses made by physicians or coders. Thus, the estimates of completeness are expected to be higher than the estimates of case coverage from the AHW-ACR record linkage after the case follow-back is complete. Results from the case follow-back should be incorporated into the results from the record linkage. Some registries use the results from the record linkage as an independent case ascertainment method to estimate completeness, without incorporating the results from the case follow-back [14, 24]. This would lead to an underestimate of registry completeness if the results from the case follow-back were similar to those results here, where approximately 66% of cases identified from the AHW data are not unregistered cancer cases.

Variations in completeness estimates existed among cancer sites investigated in the AHW-ACR record linkage. The patterns of low case coverage observed in leukemia are consistent with some studies using methods of record linkage [32, 38] but not others using methods of re-screening of cases [30] and capture/recapture [63].

The fact that female breast cancers had the highest complete coverage of registration is thought to be because of the public awareness of this common cancer site. In addition, the high proportion of biopsies being done for female breast cancer cases could have an impact on this complete coverage. Owing to screening and early detection programs, the patients with cancers of the female breast are likely to have a more complete coverage of registration. The patterns of high case coverage observed in the female breast cancers are consistent with results from some studies [32, 70], but not with others [24, 30, 38, 63, 65, 79]. Different methods used in studies associated with various case-finding mechanisms among registries should be taken into account when interpreting such variations in results.

When the comparisons were made with other studies using record linkage for completeness estimates, the use of population-based hospital discharge and day procedure files from the AHCIP in this study is considered most comprehensive, as opposed to the one by Schouten *et al* in 1993 [32] to compare the registry data to a centralized database used by general practitioners and the other by Brewster *et al* in 1997 [38] to compare the registry data to an assembled collection of databases from 14 separate sources. Thus, the results from the AHW-ACR record linkage would be considered more reliable.

2. The Incidence:Mortality (I:M) Ratio

2.1 ACR Completeness Estimates

Registration completeness estimated by the I:M ratio method can be affected by new screening and treatment programs. Changes in incidence and mortality as a result of early detection and better treatment programs for both cancers of the female breast and prostate would have impacts on I:M ratios. For melanoma of the skin, prostate, and female breast cancers, all of which have relatively long survival periods after diagnosis, I:M ratios that were much higher than 1.0 are understandable. Conversely, cancers of the pancreas and lung, which have short survival periods, had I:M ratios much close to 1.0. These I:M ratios derived from the ACR data by cancer site were similar to those published in the *CINA* by the NAACCR [72].

The overall ACR completeness estimated by I:M ratios with the inclusion of female breast and prostate cancers was the lowest among the methods used. It is thought to be because of the differences in population structure and survival experiences in the standard registry used for comparison purposes, as well as the inclusion of both female breast and prostate cancers that are excluded by the NAACCR. The NAACCR uses SEER registries and NMDB in the U.S. as the gold standard for completeness estimates, only the white population is included, a system that should be differentiated from the use of all ethnic groups in Canada. The unavailability of ethnic information in Canada is a major drawback that makes such comparisons questionable. However, the use of all races combined in completeness estimates for Canadian registries has been proposed recently [100].

In the estimation of the completeness of case ascertainment, NAACCR excludes both breast and prostate cancers. Breast cancer has increased in incidence, but decreased in mortality as a result of new screening and treatment programs. Prostate cancers have experienced dramatic changes in incidence since the introduction of Prostate Specific Antigen (PSA) in the late 1980s. However, breast and prostate cancers account for almost 30% of all invasive cancers (excluding non-melanoma skin cancer) diagnosed in Alberta. Thus, estimates of overall completeness of case ascertainment could be biased if the completeness of case ascertainment were different for the two sites. The inclusion of female breast cancer for completeness estimates has been proposed in a revised model adjusting for differences in case fatality [100].

The results of ACR completeness estimated by I:M ratios indicated that for some sites the completeness was over 100% (i.e. pancreas, lung & bronchus, and leukemia). This may be from different survival rates or trends in incidence and mortality between the ACR and the SEER registries, to random variations caused by small sample sizes, or to inaccuracies concerning the cause of death on the death certificate. For example, inaccurate coding on cause of death from Vital Statistics for female breast, prostate, and stomach cancers in the ACR would have impacts on the estimates. For cancers of the female breast and prostate, a much lower completeness estimate using I:M ratios could arise from the lower levels of screening and detection programs achieved in Canada than in the U.S.

2.2 Impacts of Different Rules on Completeness Estimates

The use of I:M ratios for completeness estimates can be affected by the different coding rules used. However, the comparison of stable I:M ratios with those of previous years in the same registry does not seem to present a significant problem unless the adoption of coding rules has been changed. In the certification of the NAACCR member registries, the SEER registries, together with the NMDB in the U.S., are used as the gold standard. Concern has been expressed over the effect that the differences in coding rules would have on the final estimates because some member registries use IARC or other rules for coding. The standardization of coding rules in the NAACCR could make the data comparisons more meaningful among cancer registries in North America. In particular, the use of I:M ratios for completeness estimates can be enhanced by the standardization of coding rules among registries. Lower cancer incidence and I:M ratios would be expected when IARC/IACR rules are adopted because of the time independence on coding multiple primaries.

The discrepancies resulting from the adoption of two rules did not indicate marked differences in completeness estimates. However, registries may be penalized for their choice of coding rules if the completeness of IARC registries is close to a critical level in the NAACCR certification process.

3. Percent Histological Verification (HV%) and Percent Death Certificate Only (DCO%) as Completeness Indicators

The results of using DCO% for registration completeness estimates were consistent with the results found when using HV% as a completeness indicator. In general, cancer

registries with more cases confirmed histologically should have fewer cases diagnosed by DCO. However, HV% may not be a reliable indicator for completeness estimates for registries using pathology reports as a main source for registering cases as in the ACR.

When using HV% as a completeness indicator, patients having cancers such as pancreatic cancer with shorter survival periods after diagnosis would have lower HV%. This is so because invasive treatment or confirmation may not be considered appropriate for patients with such a low survival period. Patients with lung cancer also tended to have a slightly lower percentage of cases histologically verified as opposed to those with cancers with longer survival periods such as female breast and prostate cancers. This could be explained by the fact that the survival period for lung cancer patients is relatively short.

4. **Case Re-finding in the UAH**

The number of records in the UAH data set was used as the denominator for ACR completeness estimates under the UAH case re-finding exercise. This was chosen instead of the number of records in the ACR data as used as the denominator in the AHW-ACR record linkage. The UAH data included only a small proportion of cancer cases diagnosed in Alberta. Thus using the number of records in the UAH data as the denominator seems appropriate. Thus, the results from the AHW-ACR record linkage should be considered more reliable than those from the UAH case re-finding. Unlike the results from the AHW-ACR linkage, leukemia had the highest degree of case coverage in the UAH case re-finding exercise. The pattern that patients with pancreatic cancer had the

lowest case coverage in the UAH case re-finding was consistent with the results from using indicators of HV% and DCO%.

In addition, the different sources of data between the AHW-ACR record linkage and UAH case re-finding should be recognized when interpreting the results. Inpatient data were used in the AHW-ACR record linkage, and outpatient data were used under the UAH case re-finding method.

B. Stratification by Demographic and Geographic Factors

1. **Sex Stratification**

Females had higher case coverage overall with all sites included when results from the AHW-ACR record linkage were stratified by sex. Conversely, females had lower case coverage overall with all sites included using the methods of I:M ratios and UAH case re-finding. These patterns were not consistent across the sites within methods and were also observed in the study by Schouten *et al* in 1993 [32]. Owing to the fact that females are more likely to seek medical attention whenever they have problems with their health, they are more likely to enter the health care system and have more complete coverage in registration. The pattern that females have higher case coverage than males was consistent with the study by Navarro *et al* in 1986, for cancers of colon-rectum and pancreas, but not for lung cancer [73].

2. **Age Group Stratification**

Variations in completeness estimates existed among cancer sites when comparisons were made by age group. Thus, the pattern of results stratified by age group should be interpreted on a site-specific basis. Overall, the degree of completeness was higher in the middle age groups and was lower in the oldest groups across methods used. Completeness estimates for the youngest age group were not stable because of the small sample size in that particular age group. The pattern that the oldest group had lower case coverage was because older people are more likely to have more health problems other than cancers, and that they may not undergo extensive investigations for their cancers. Besides, it was found that more cases without histological confirmation across all major cancer sites were in the older age groups. This can explain the lower case coverage observed in older age groups. The trends in age group comparisons observed are consistent with results from some studies using record linkage and capture/recapture methods [32, 65, 69], but not in others using statistical techniques [79].

3. **Regional Stratification**

The pattern that registration completeness was higher in urban than in rural regions was consistent for most of the sites investigated across methods except for the I:M ratios. This pattern verified the assumption we had initially, namely that more completeness would have been achieved in urban than in rural regions because of urban regions having well-developed electronic databases, more convenient transportation, and better communication and registry automation systems than those in rural regions. This pattern is consistent with the study by Larsson in 1971 using indicators of HV% [58]. However,

the pattern that urban regions had lower registration completeness than rural regions for most sites investigated when using I:M ratios is not easily explained and should be further investigated.

C. Comparison of Methods

When comparisons are made of the ACR completeness estimates by comparing the results from the AHW-ACR record linkage, the I:M ratio method, and indicators of HV% and DCO%, congruency and incongruency are found in various respects. Among all of the independent case ascertainment methods used, the results from the AHW-ACR record linkage are considered the most reliable because ACR data were compared to the most comprehensive case coverage file in Alberta, that maintained by AHW. Concerns have been expressed over the use of SEER and U.S. mortality as the standard registry to derive completeness estimates by the I:M ratio method for Canadian registries because of the different population structure and survival experience existing in the two countries. The choice of the UAH as the only hospital for case re-finding may lead to biased estimates, because the UAH covers only a small proportion of all cancer cases diagnosed in Alberta. The results from the use of other hospitals may be quite different.

In addition to the reliability of the results, the criteria for the choice of the most appropriate methods should depend on the costs and resources associated with them, along with the degree of ease of access among registries. In terms of costs and resources associated with methods, I:M ratios as absolute measures and indicators of HV% and DCO% as relative measures appear to be simple and inexpensive to use. However, results

from indicators of HV% seem not to be as reliable because pathology reports are one of the main sources that the ACR uses to register cancer cases.

The record linkage between AHW data and ACR data for ACR completeness estimates appears to be the most appropriate based on the case reporting mechanism in Alberta and the resources associated with the linkage. The reliability of the results is one of the major advantages. However, the expertise required for conducting the computerized record linkage, as well as the effort and costs needed to implement case follow-back should be recognized. Owing to various advantages and features that different methods have, one method should be used to supplement another as suggested in the literature.

D. Strengths and Limitations of The Current Study

1. **Strengths**

The contribution of cancer registry data to cancer control is critical. It includes the use of registry data in the surveillance of cancer in the population, the evaluation of cancer control programs, as well as the use of registry data in epidemiological studies to investigate risk factors. This study can contribute to overall cancer control in that the data quality in cancer registries can be assessed through the methods used in this study. Better planning and quality assurance programs can be implemented to improve cancer registration through the factors identified that contribute to incompleteness.

The completeness of case ascertainment in the ACR was estimated using the various methods as described. The main method used in this study was a comparison of

ACR data to independent data -- Hospital Discharge and Day Procedure data from AHW. A deterministic approach was used in this record linkage because of the existence of unique identifying information – Personal Heath Number (PHN) between these two data sets. Minimal omissions and errors in the data sets are other reasons that deterministic record linkage was used. The popularity for the use of record linkage in epidemiological studies has been shown to be both efficient and cost-effective.

The existence of a provincial health care plan also made this study feasible. All health services provided to inpatient cancer cases in Alberta have records with AHW. These records are an invaluable source for research in health care utilization, program evaluation, and policy recommendations, as well as in various epidemiological studies, including case-control studies, cohort studies, and descriptive studies.

Furthermore, the use of the I:M ratio method and indicators of HV% and DCO% derived from the ACR data can help us understand the mechanisms associated with registration completeness. When these indices were compared with those derived from registries with similar registration mechanisms or with patterns observed previously, consistency and inconsistency can be illustrated. Therefore, it is useful to consider the results from these indices together with the results derived from other methods.

2. **Limitations**

The main method used in this study was a comparison of ACR data to Hospital Discharge and Day Procedure data from AHW. Although a deterministic approach was adopted in the record linkage because of the coexistence of unique identifiers in both data sets, the results may have been be different had a probabilistic approach been used. It

might not have been appropriate to adopt a deterministic approach when the data quality in the data sets was not assessed. Errors and omissions in data sets can result in fewer matched cases in a deterministic linkage than in a probabilistic linkage. Despite the fact that data quality in the ACR has been determined in previous studies, a comparison of results between the two approaches used would have reflected discrepancies in matching identical cases. The results from using probabilistic linkage should be available to allow comparisons to find discrepancies, if any [42].

Further, we are still waiting for responses from hospitals for some of the questionable cases requiring follow-back. Owing to the slowness of this process, only the most recent results of cases with diagnoses confirmed can be incorporated into the results from the record linkage. Results from case follow-back indicate that the overall completeness estimates will be higher than the case coverage results we currently have.

The use of the SEER incidence and the NMDB mortality in the U.S. as a standard registry to estimate ACR completeness through the method of I:M ratios may not be appropriate for all Canadian registries because of the differences in survival experiences and coding rules used for multiple primaries. In addition, the inclusions of only white populations with a higher social class in their data and the data with all races combined in Canada may make the comparisons questionable. The fact that different levels of screening programs achieved for both female breast and prostate cancers and different health care systems between two countries should also be taken into account.

Improvements could be made by using indicators of DCN% rather than DCO% for completeness estimates in the study, because DCN% is a better indicator than DCO% as suggested in the literature [10].

The results of the case re-finding exercise in the UAH may be biased because the choice of hospitals should have been random based on the different caseloads and the locations of the hospitals. Owing to the advantages of the easy access to and cost-efficiencies of the UAH, it was chosen, and the results were incorporated to supplement the methods used in the AHW-ACR linkage. It is quite possible that contradictory results could have been obtained if hospitals other than the UAH had been chosen.

CHAPTER VII. CONCLUSIONS and RECOMMENDATIONS

The primary objectives of this study were to determine the registration completeness of the ACR, to investigate factors contributing to its completeness, and to compare methods used in the completeness estimates. Overall, the completeness of the ACR was good for the years 1994-96, but with variations existing across cancer sites, demographic factors such as sex and age, and geographic factors such as place of residence. The results of the completeness estimates of this study were very close to those achieved by the NAACCR using the I:M ratio method. Variations in completeness estimates among cancer sites were consistent with results from some studies using either different or the same methods. Variations of completeness estimates in demographic and geographic factors need to be further investigated for improvement of cancer registration. Method comparisons indicated that the AHW-ACR record linkage was the most appropriate method used, along with the methods of I:M ratios and indicators of HV% and DCO%. These provided indications of ACR completeness, based on its case-finding mechanisms and registration procedures. Results from and the experience of this study may be used to contribute to the national case ascertainment study in Canada, and can be generalized to other registries to improve cancer registration for the better control of cancer.

The recommendations from this study are as follows:

1. Both deterministic and probabilistic approaches should be implemented for the AHW-ACR record linkage to compare the results and to investigate the possible discrepancies. One approach should be used to supplement the other.
2. The selection of hospitals for case re-finding exercises should be unbiased. Hospitals should be chosen randomly based on caseloads and geographic locations.
3. Statistical techniques can be used to model registration completeness by incorporating various factors associated with completeness simultaneously for a better understanding and improvement of cancer registration.
4. Improvements of cancer registration in Alberta can be achieved through comparing the ACR data to AHW data on a routine basis. Results from this study indicate that emphasis should be placed on the groups with lower case coverage, including people in the older age groups, people living in the rural regions, and patients with leukemia.

REFERENCES

1. IARC Scientific Publications No.21 (1978): *Cancer Registration and Its Techniques.* Lyon, France: International Agency for Research on Cancer.

2. IARC Scientific Publications No.66 (1985): *The Role of The Registry in Cancer Control.* Lyon, France: International Agency for Research on Cancer.

3. IARC Scientific Publications No.95 (1991): *Cancer Registration: Principles and Methods*. Lyon, France: International Agency for Research on Cancer.

4. Hutchison CL, Roffers SD., Fritz AG. (1997): *Cancer Registry Management: Principles & Practice.* Lenexa, Kansas, U.S.A.: National Cancer Registrars Association, Inc.

5. Bender AP, Jagger HG, Fraser J, et al. Feasibility study of a statewide pathology-based cancer surveillance system in Minnesota. J Med Systems 1987;11(1):25-44.

6. Menck H, Smart C (1994). Central Cancer Registries: Design, Management, and Use. Switzerland: Harwood Academic Publishers GmbH.

7. World Health Organization (1997): *Biennial Report 1996/1997*. Lyon, France: International Agency for Research on Cancer.

8. Brewster D, Muir C, Crichton J. Registration of colorectal cancer in Scotland: an assessment of data accuracy based on review of medical records. Public Health 1995; 109: 285-292.

9. The Ontario Cancer Registry (1995): *Cancer Incidence in Ontario: Trends and Regional Variations*. The Ontario Cancer Treatment and Research Foundation.

10. IARC Technical Report No.19 (1994): *Comparability and Quality Control in Cancer Registration*. Lyon, France: International Agency for Research on Cancer.

11. Kuntoro, LaPorte RE, Mazumdar S. Approaches to quality control with an application to a new cancer registry in a developing country. J Clin Epidemiol 1994;47(7):779-786.

12. Hawkins MM., Swerdlow AJ. Completeness of cancer and death follow-up obtained through the National Health Service Central Register for England and Wales. Br J Cancer 1992; 66:408-413.

13. Murray CL, Rose MA, Swenson KK, et al. Evaluation of oncology registry follow-up methods. Cancer 1995; 76:880-9.

14. Kardara M, Acquilla S, Forster D, et al. Establishing baseline data in cancer registration in northern England: implications for Health of the Nation targets. J Epidemiol Community Health 1995; 49:150-152.

15. Mattsson B, Wallgren A. Completeness of the Swedish Cancer Register. Non-notified cancer cases recorded on death certificates in 1978. Acta Radiologica-Oncology 1984;23(5):305-13.

16. Warnakulasuriya KAAS, Acworth P, Bell J, et al. Incompleteness of oral cancer registration in south-east England, 1971-87. Br J Cancer 1994;70:736-738.

17. Mork J, Thoresen S, Faye-Lund H, et al. Head and neck cancer in Norway. A study of the quality of the Cancer Registry of Norway's data on head and neck cancer for the period 1953-1991. APMIS 1995;103(5):375-82.

18. Mukherjee AK, Leck I, Langley FA, et al. The completeness and accuracy of health authority and cancer registry records according to a study of ovarian neoplasms. Public Health 1991;105(1):69-78.

19. Harvei S, Tretli S, Langmark F. Quality of prostate cancer data in the cancer registry of Norway. Eur J Cancer 1996;32A(1):104-110.

20. Silcocks PBS, Thornton-Jones H, Skeet RG. Can we achieve 100% ascertainment in cancer registration? Public Health 1989;103:23-30.

21. Fleming ID, Phillips JL, Menck HR. The National Cancer Data Base report on completeness of American Joint Committee on cancer staging in United States cancer facilities. Cancer 1996;78(7):1498-504.

22. Mettlin CJ, Menck HR, Winchester DP, et al. A comparison of breast, colorectal, lung, and prostate cancers reported to the National Cancer Data Base and the Surveillance, Epidemiology, and End Results program. Cancer 1997;79(10):2052-61.

23. Benn RT, Leck I, Nwene UP. Estimation of completeness of cancer registration. Intl J Epidemiol 1982;11(4):362-367.

24. Nwene U, Smith A. Assessing completeness of cancer registration in the north-western region of England by a method of independent comparison. Br J Cancer 1982;46:635-639.

25. Berkel J. General practitioners and completeness of cancer registry. J Epidemiol Community Health 1990;44:121-124.

26. Melia J, Frost T, Graham-Brown R, et al. Problems with registration of cutaneous malignant melanoma in England. Br J Cancer 1995;72:224-228.

27. Garne JP, Aspegren K, Moller T. Validity of breast cancer registration from one hospital into the Swedish National Cancer Registry 1971-1991. Acta Oncologica 1995;34(2):153-156.

28. Schouten LJ, Does-van den Berg AV, Otter R, et al. Accuracy and completeness of the registration of childhood leukaemia in the Netherlands, 1989-1992. Eur J Cancer 1997;33(6):891-894.

29. Swerdlow AJ, Douglas AJ, Vaughan Hudson G, et al. Completeness of cancer registration in England and Wales: an assessment based on 2,145 patients with Hodgkin's disease independently registered by the British National Lymphoma Investigation. Br J Cancer 1993;67:326-329.

30. Zippin C, Lum D, Hankey BF. Completeness of hospital cancer case reporting from the SEER program of the National Cancer Institute. Cancer 1995;76(11):2343-50.

31. Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Acta Oncologica 1994;33(4):365-369.

32. Schouten LJ, Hoppener P, Van Den Brandt PA, et al. Completeness of cancer registration in Limburg, the Netherlands. Intl J Epidemiol 1993;22(3):369-376.

33. Nielsen GL, Sorensen HT, Pedersen AB, et al. Analysis of data quality in registries concerning diabetes mellitus – a comparison between a population based hospital discharge and an insulin prescription registry. J Med Systems 1996;20(1):1-10.

34. Kyllonen LE, Teppo L, Lehtonen M. Completeness and accuracy of registration of colorectal cancer in Finland. Ann Chirurgiae et Gynaecologiae 1987;76(4):185-90.

35. Flam F, Rutqvist LE. Under-registration of gestational trophoblastic disease in the Swedish Cancer Registry. Eur J Epidemiol 1992;8(5):683-6.

36. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Intl J Epidemiol 1996;25(2):435-42.

37. Storm HH. Completeness of cancer registration in Denmark 1943-1966 and efficacy of record linkage procedures. Intl J Epidemiol 1988;17(1):44-9.

38. Brewster DH, Crichton J, Harvey JC, et al. Completeness of case ascertainment in a Scottish regional cancer registry for the year 1992. Public Health 1997;111(5):339-43.

39. Zanetti R, Vigano C, De Molli S, et al. Comparative completeness and correspondence of cancer mortality data as collected by ISTAT and Cancer Registries. Tumori 1982;68(6):457-63.

40. Goldberg J, Gelfand, HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Rev 1980; 2:210-220.

41. IARC Scientific Publications No.15 (1976): *Cancer Incidence in Five Continents*. Lyon, France: International Agency for Research on Cancer.

42. Turner DCR. Cholecystectomy as a factor for colon cancer. Doctoral Thesis 1997.

43. Newcombe HB and Kennedy JM. Record Linkage. Communications of the Association for Computing Machinery 1962;5:563-566.

44. Fellegi IP and Sunter AB. A theory for record linkage. Journal of the American Statistical Association 1969;64:1183-1210.

45. Jaro MA. Probabilistic linkage of large public health data files. Statistics in Medicine 1995;14:491-498.

46. Roos LL and Wajda A. Record linkage strategies: Part I. Estimating information and evaluating approaches. Meth Inform Med 1991;30:117-23.

47. Newcombe HB, Fair ME, Lalonde P. Discriminating powers of partial agreements of names for linking personal records. Part I. The logical basis. Meth Inform Med 1989;28:86-91.

48. Newcombe HB, Fair ME, Lalonde P. Discriminating powers of partial agreements of names for linking personal records. Part II. The empirical test. Meth Inform Med 1989;28:92-6.

49. Fenna D. Phonetic reduction of names. Computer Programs in Biomedicine 1984; 19(1):31-6.

50. Goehring R. Identification of patients in medical data bases – soundex codes versus match code. Medical Informatics 1985;10(1):27-34.

51. Davis KB, Fisher L, Gillespie MJ, Pettinger M. A test of the National Death Index using the Coronary Artery Surgery Study (CASS). Controlled Clinical Trials 1985;6(3):179-91.

52. Sideli RV, Friedman C. Validating patient names in an integrated clinical information system. Proceedings – the Annual Symposium on Computer Applications in Medical Care 1991;588-92.

53. Mortimer JY, Salathiel JA. 'Soundex' codes of surnames provide confidentiality and accuracy in a national HIV database. Communicable Disease Report, CDR Review 1995;5(12):R183-6.

54. Balogun MA, Wall PG, Noone A. Undernotification of tuberculosis in patients with AIDS. International Journal of STD & AIDS 1996;7(1):58-60.

55. Seaman SR, Brettle RP, Gore SM. Mortality from overdose among injecting drug users recently released from prison: database linkage study. Br J Cancer 1998;316(7129):426-8.

56. Cooksley CD, Hwang LY, Waller DK, Ford CE. HIV-related malignancies: community-based study using linkage of cancer registry and HIV registry data. International Journal of STD & AIDS 1999;10(12):795-802.

57. Alberta Cancer Registry: *Alberta Cancer Registry Coding Manual*. May 1999.

58. Larsson S. Completeness and reliability of lung cancer registration in the Sweden Cancer Registry. Acta Path Microbiol Scand Section A 1971;79:389-398.

59. Driver CR, Braden CR, Nieves RL, et al. Completeness of tuberculosis case reporting, San Juan and Caguas Regions, Puerto Rico, 1992. Public Health Reports 1996;111(2):157-61.

60. Heiberger RM, Miller CL, Feigl P, et al. A novel method of assessing completeness of tumor registration. Cancer 1983;51(12):2362-6.

61. Brewster D, Muir C, Crichton J. Registration of non-melanoma skin cancers in Scotland – how accurate are site and morphology codes? Clin Exp Dermatology 1995;20(5):401-5.

62. Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. J Chron Dis 1974;27:25-36.

63. Robles SC, Marrett LD, Clarke EA, Risch HA. An application of capture-recapture methods to the estimation of completeness of cancer registration. J Clin Epidemiol 1988;41(5):495-501.

64. Hook EB and Regal RR. The value of capture-recapture methods even for apparent exhaustive surveys: the need for adjustment for source of ascertainment intersection in attempted complete prevalence studies. Am J Epidemiol 1992;135(9):1060-67.

65. Brenner H, Stegmaier C, Ziegler H. Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. J Epidemiol Community Health 1995;49:426-430.

66. Hook EB and Regal RR. Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence. Am J Epidemiol 1995;142:S48-S52.

67. Hook EB and Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiologic Review 1995;17(2):243-264.

68. Cormack RM. The statistics of capture-recapture methods. Oceanogr. March Biol Ann Rev 1968;6:455-506.

69. Brenner H, Stegmaier C, Ziegler H. Estimating completeness of cancer registration in Saarland/Germany with capture-recapture methods. Eur J Cancer 1994;30A(11):1659-1663.

70. Schouten LJ, Straatman H, Kiemeney LA, et al. The capture-recapture method for estimation of cancer registry completeness: a useful tool? Intl J Epidemiol 1994;23(6):1111-6.

71. Kim JP, Park IS, Ahn YO, et al. 1991 cancer incidence in Seoul, Korea: results of the implementation study of the Seoul Cancer Registry. Journal of Korean Medical Science 1995;10(2):74-84.

72. NAACCR Publication. (1999): *Cancer in North America, 1991-1995.* North American Association of Central Cancer Registries (NAACCR).

73. Navarro C, Perez-Flores D, Coleman MP. Cancer incidence in Murcia, Spain, in 1982: first results from a population-based cancer registry. Int J Cancer 1986; 38:1-7.

74. Villard-Mackintosh L, Coleman MP, Vessey MP. The completeness of cancer registration in England: an assessment from the Oxford-FPA contraceptive study. Br J Cancer 1988; 58: 507-511.

75. Counsell CE, Collie DA, Grant R. Limitations of using a cancer registry to identify incident primary intracranial tumours. Journal of Neurology, Neurosurgery & Psychiatry 1997;63(1):94-7.

76. Capocaccia R, Angelis RD. Estimating the completeness of prevalence based on cancer registry data. Statistics in Medicine 1997; 16:425-440.

77. van der Sanden GAC, Coebergh JWW., Schouten LJ, et al. Cancer incidence in the Netherlands in 1989 and 1990: first results of the Nationwide Netherlands Cancer Registry. Eur J Cancer 1995; 31A(11):1822-1829.

78. Hilsenbeck SG, Kurucz C, Duncan RC. Estimation of completeness and adjustment of age-specific and age-standardized incidence rates. Biometrics 1992; 48:1249-1262.

79. Bullard J, Coleman MP, Robinson D, et al. Completeness of cancer registration: a new method for routine use. Br J Cancer 2000;82(5):1111-6.

80. Canadian Cancer Registries Data Quality Committee (1999): *Overview of cancer registration in Canada*. Ottawa. "Maximizing our potential workshop".

81. *Canadian Cancer Statistics 2000*. Ontario, Canada: National Cancer Institute of Canada.

82. Quality Management Working Group (1998): *Report of the Working Group on Quality Management*. Canadian Coalition on Cancer Surveillance (CCOCS).

83. O'Sullivan B, Gospodarowicz M, Mackillop WJ, et al. The consultation to develop a national strategy for cancer staging in Canada. Cancer Prevention & Control 1998;2(6):287-94.

84. Liu CF, Soskolne CL, Hatcher J. The effect of different coding rules on estimates of registry completeness. J of Registry Management 2001 (submitted).

85. IARC Scientific Publications No. 78 (1986): *Carcinogenicity of alkylating cytostatic drugs: conclusions and directions for future research*. Lyon, France: International Agency for Research on Cancer.

86. Boice JD Jr, Greene MH, Killen JY, et al. Leukemia and preleukemia after adjuvant treatment of gastrointestinal cancer with semustine (methyl-CCNU). N Engl J Med 1983;309(18):1079-84.

87. Hankey BF, Curtis RE, Naughton MD, et al. A retrospective cohort analysis of second breast cancer risk for primary breast cancer patients with an assessment of the effect of radiation therapy. J Natl Cancer Inst 1983;70(5):797-804.

88. Pelham JM, Gray JD, Flannery GR, et al. Interferon-alpha conjugation to human osteogenic sarcoma monoclonal antibody 791T/36. Cancer Immunology, Immunotherapy 1983;15(3):210-6.

89. Mann BD, Storm FK, Morton DL, et al. Predictability of response to clinical thermochemotherapy by the clonogenic assay. Cancer 1983;52(8):1389-94.

90. Coughlin SS and Beauchamp TL (1996). *Ethics and Epidemiology*. New York/Oxford: Oxford University Press.

91. *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.* Acts of Parliament, Ottawa 1999.

92. Beauchamp TL, Childress JF (1994). Principles of biomedical ethics (4th ed.). New York:Oxford University Press.

93. Canadian Institute for Health Information (CIHI) (1999). *Privacy and Confidentiality of Health Information at CIHI: principles and policies for the protection of health information.*

94. Canadian Institute for Health Information (CIHI) (1995). *Health Data Sharing in Canada: a resource guide.*

95. Province of Alberta: *Health Information Act*. Revised Statutes of Alberta 2000.

96. Province of Alberta: *Cancer Programs Act*. Revised Statutes of Alberta 1999.

97. Legislative Working Group (1999): *Report of the Working Group on Privacy of Health Information*. Canadian Coalition on Cancer Surveillance (CCOCS).

98. NAACCR Annual Meeting Poster Presentation (1994): *Comparison of automated with manual cancer data capture from hospital discharge summaries over a two-year period in Alberta, Canada*. Division of Epidemiology & Preventive Oncology, Alberta Cancer Board.

99. Brenner H. Limitations of the death certificate only index as a measure of incompleteness of cancer registration. Br J Cancer 1995; 72(2):506-510.

100. Tucker TC, Howe HL. Measuring the quality of population-based cancer registries: the NAACCR perspective. J Reg Mgt 2001;28(1):41-44.